DOCUMENT RESUME

ED 127 806 FL 007 948

AUTHOR	Shapiro, Michael C.; Schiffman, Harold F.
TITLE	Language and Society in South Asia. Final Report.
INSTITUTION	Institute of International Studies (DHEW/OE), Washington, D.C.
BUREAU NO	BR-N0012HH
PUB DATE	Sep 75
CONTRACT	OEC-0-74-2093
NOTE	380p.
EDRS PRICE	MF-$0.83 HC-$20.75 Plus Postage.
DESCRIPTORS	*Asian Studies; *Bilingualism; Burmese; Cultural Context; *Dialects; Dialect Studies; Dravidian Languages; Language Classification; *Language Variation; Linguistic Borrowing; Multilingualism; Regional Dialects; Social Dialects; *Sociolinguistics; Tibetan
IDENTIFIERS	*Asia (South); *Code Switching; Indo Aryan Languages; Munda Languages; Tibeto Burman Languages

ABSTRACT

This work attempts to provide an overview of linguistic diversity in South Asia and to place this diversity in a cultural context. The work tries to describe the current state of knowledge concerning socially conditioned language variation in the subcontinent. Each of five major language families contains numerous mutually intelligible and unintelligible dialects. Different dialects of a language may be required for written and spoken use and for different social groups. Bilingualism and multilingualism are common for communication between groups. Language choice is important for education, politics, radio and television. Chapter 2 of this book enumerates criteria used in the taxonomy of language forms, discussing a number of theories of dialect formation from the points of view of linguistic innovation and diffusion of linguistic change. Chapter 3 surveys literature on classification of South Asian languages. Chapter 4 considers South Asia as a distinct linguistic area and Chapter 5 evaluates literature on South Asian social dialects. Chapter 6 examines linguistic codes encompassing elements from more than one autonomous language. Chapter 7 considers the ways in which the lexicon of South Asian languages and dialects contain elements that structure themselves into concrete systems. (CHK)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED127806

TITLE VI, SEC. 602, NDEA
Contract No. 0-74-2093
Project No. N0012HH

final report FL

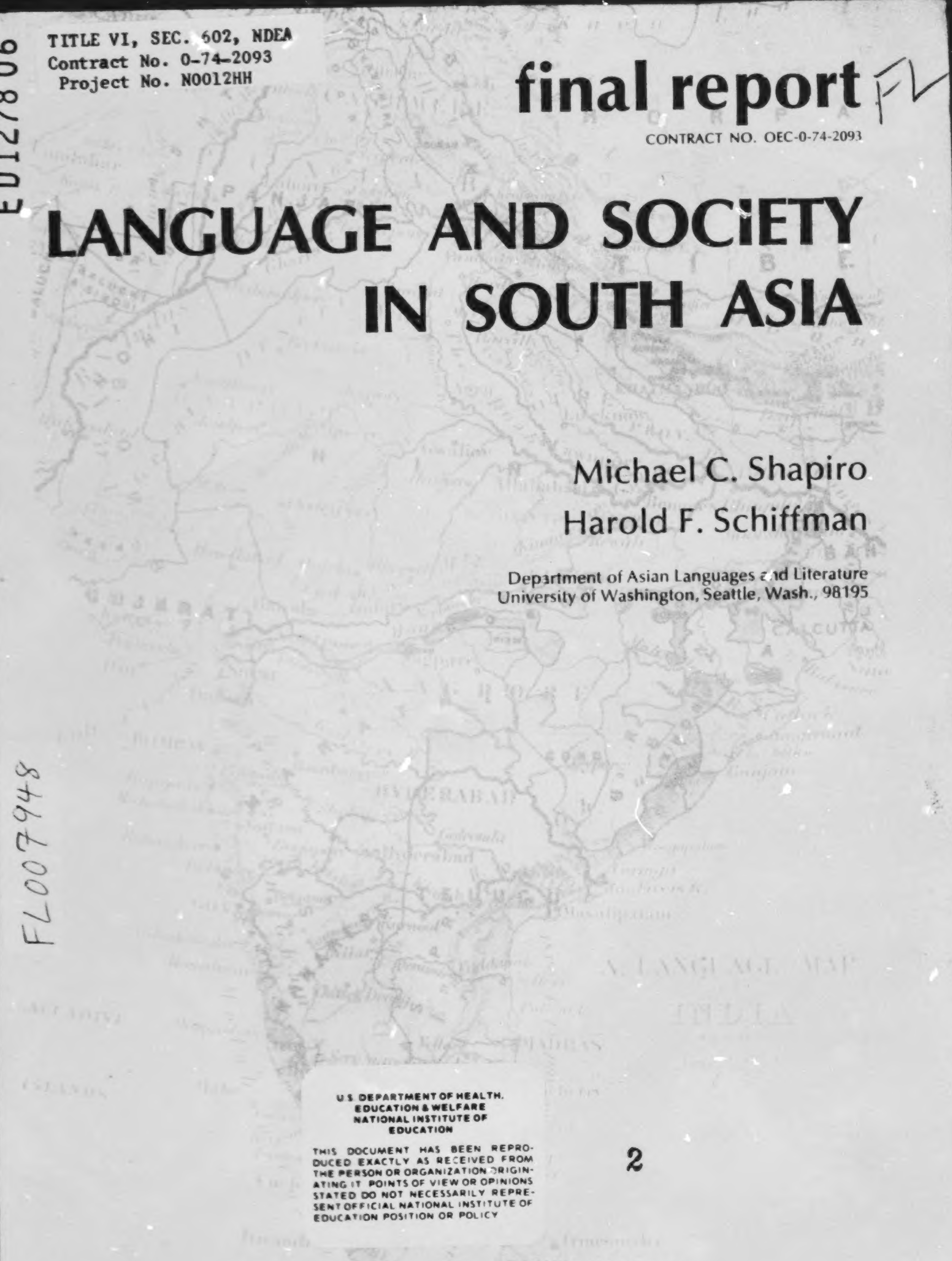
CONTRACT NO. OEC-0-74-2093

LANGUAGE AND SOCIETY IN SOUTH ASIA

Michael C. Shapiro
Harold F. Schiffman

Department of Asian Languages and Literature
University of Washington, Seattle, Wash., 98195

FL007948



U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

September, 1975, U.S. Department of Health Education and Welfare, Office of Education, International Studies Branch

LANGUAGE AND SOCIETY IN SOUTH ASIA

Michael C. Shapiro

Harold F. Schiffman

The research reported herein
was performed pursuant to a
contract with the United States
Office of Education, Department
of Health, Education, and
Welfare (Institute of International
Studies), under the authority of
Section 602, NDEA.
1975

TABLE OF CONTENTS

Preface and Acknowledgements viii

1. Introduction. 1

2. The Theoretical Description of Language and Dialect 23
   - 2.0. Introduction. 23
   - 2.1. Isoglossal Theories of Language and Dialect 24
   - 2.2. *Stammbaum* and Wave Models of Linguistic History 31
   - 2.3. Structural Dialectological Models of Linguistic Relatedness 46
   - 2.4. Transformational Analyses of Dialect Differences. 56
   - 2.5. Sociolinguistic Models of Language and Dialect. 64
   - 2.6. The Linguistic Basis of Typological Studies of South Asian Languages 75
   - Notes 87

3. Traditional Taxonomies of South Asian Languages. 97
   - 3.0. Introduction. 97
   - 3.1. Classification of the Indo-Aryan Languages. 100
   - 3.2. Classification of the Dravidian Languages 117
   - 3.3. Classification of the Munda Languages 137
   - 3.4. Classification of the Tibeto-Burman Languages 139
   - Notes 149

4. South Asia as a Linguistic Area 153
   - 4.0. Introduction. 153
   - 4.1. Linguistic Basis of the South Asian Linguistic Area . . .154
   - 4.2. Micro-Linguistic Area Phenomena 172
   - 4.3. Historical Implications of a South Asian Linguistic Area. 179
   - 4.4. Mechanisms for Genesis of a South Asian Linguistic Area .184
   - 4.5. Conclusions and Desiderata. 187
   - Notes 190

5. Social Dialectology 195
   - 5.0. Introduction. 195
   - 5.1. General Problems of Social Dialectology in South Asia . .196
   - 5.2. Studies of Caste Dialects 198
   - 5.3. Social Variables other than Caste 207

5.4. Diglossia. 211

5.5. Cross Language Variability and Change. 222

5.6. Other Sociolinguistic Topics 223

5.7. Conclusions and Desiderata 225

Notes. 227

6. Individual and Group Linguistic Repertoires. 229

6.0. Bi- and Multilingualism in South Asia. 229

6.1. Census Data on South Asian Bi- and Multilingualism 230

6.2. General and Theoretical Studies of Bilingualism. 232

6.3. Studies of Bilingualism in South Asia. 235

6.4. Pidginization and Creolization 246

6.5. South Asian English and the Influence of English on South Asian Languages. 269

Notes 280

7. Ethnographic Semantics and the Ethnography of Speaking 285

7.0. Introduction and Theoretical Discussion. 285

7.1. Ethnographic Semantics 287

7.1.1 Kinship 287

7.1.2. Structural Approaches to the Study of Kinship in South Asia. 292

7.1.3. Data-Oriented Systems of a Non-Formal Type 295

7.1.4. Formal Analysis Studies Proper 296

7.1.5. Applied and Comparative Studies. 302

7.1.6. Fictive Kinship. 303

7.2. The Ethnography of Speaking. 307

7.2.0. Introduction. 307

7.2.1. Aspects of the Ethnography of Speaking 309

7.2.2. Literature on the Ethnography of Speaking in South Asia. 313

Notes. 333

Bibliography 345

LIST OF FIGURES

Page

Figure 1. Alternate Forms for *cottage cheese*. 25
Figure 2. Isoglosses in ten Swabian Villages. 26
Figure 3. Phonetic differences in the Pronunciation of ten German Lexical Items. 27
Figure 4. France: Outer Limits of Northern Phonological Features. . 28
Figure 5. Bundles of Heterolexes in the Great Lakes Area. 29
Figure 6. Speech Areas of the Eastern States. 30
Figure 7. Classification of Munda Languages 33
Figure 8. Stammbaum of Indo-Aryan Languages 34
Figure 9. The Munda Languages Arranged Typologically. 35
Figure 10. Stammbaum of Indo-European Languages [version 1]. 36
Figure 11. Reconstruction of Primitive Indonesian Consonants 37
Figure 12. Stammbaum of the Indo-European Languages [version 2]. . . 39
Figure 13. "Overlapping" Diagram of Indo-European Family of Languages. 40
Figure 14. Germanic Consonant Shifts. 43
Figure 15. Map of "Rhenish Fan" 43
Figure 16. Wave Spread in Time and Space. 45
Figure 17 Wave with Spread Arrested in One Direction 46
Figure 18. The Vowel in *man* in Language X 50
Figure 19. Structural Analysis of Dialect Structure of Language X Based on Vowel in *man* 51
Figure 20. Phonetic Boundaries of Northeastern Poland 52
Figure 21. Phonemic Areas of Northeastern Poland. 53
Figure 22. Schema of "Aspects" Models of Language Structure 59
Figure 23. Plural in Three Spanish Dialects 61
Figure 24. Class Stratification Diagram for *r* 69
Figure 25 Linguistic Effects in Frequency of Final Consonant Cluster Simplification in Detroit Black Speech 71
Figure 26. Hypothetical Ordering of Three Constraints on a Variable Rule. 71
Figure 27. The Indo-Aryan Languages (from Beames) 103
Figure 28. The Development of the Aryan Speech in India (Chatterji after Grierson) 113

Figure 29. Cardona Summary. 115
Figure 30. Classification of the Dravidian Languages (Grierson after Caldwell). .118
Figure 31. Classification of the Dravidian Languages (after Andronov). 121
Figure 32. Classification of the Dravidian Languages (Andronov 2) .122
Figure 33. The Central Dravidian Languages (after Subrahmanyam) . .123
Figure 34. The South Dravidian Languages (from Emeneau 1967). 124
Figure 35. Three-Dimensional Model of South Dravidian (from Emeneau 1967). .125
Figure 36. The South Dravidian Languages (Zvelebil) 126
Figure 37. The South Dravidian Languages (Krishnamurti) 127
Figure 38. The Central Dravidian Languages (Krishnamurti) 127
Figure 39. The North Dravidian Languages (Krishnamurti) 128
Figure 40. The Dravidian Languages (Zvelebil) 128
Figure 41. Stammbaum of the Dravidian Languages (Subrahmanyam). . .130
Figure 42. The Munda Languages (Zide) 140
Figure 43. The Baric Division (Shafer). 142
Figure 44. The Burmic Division (after Shafer,1955). 143
Figure 45. The Kukish Section of Burmic (after Shafer 1955) 144
Figure 46. The Bodic Division (after Shafer 1955) 146
Figure 47. The Bodish Section of Bodic (after Shafer 1955). 147
Figure 48. The West Himalayish Section of Bodic (after Shafer 1955)146
Figure 49. The East Himalayish Section of Bodic (after Shafer 1955)148
Figure 50. The "Burmese" Languages (after Miller 1969). 148
Figure 51. Five Functions of Sanskrit *api* in Four NIA Languages (from Emeneau 1974a) 167
Figure 52. Distribution of Typological Features (from Masica 1971).173
Figure 53. Nepali and Newari Verb Tense Forms (from Bendix 1974). .175
Figure 54. Approximate Tense-Aspect Correspondences in Newari and Nepali (from Bendix 1974). 175
Figure 55. Phonological Correspondences in Literary and Colloquial Tamil (Shanmugam Pillai 1960) 213
Figure 56. CB-SB Lexical Differences (Dimock 1960:47). 218
Figure 57. Differences in Bengali Noun Endings (Dimock 1960:49) . .218
Figure 58. Evolution of Sinhala from Sanskrit (DeSilva, 1974) . . 220
Figure 59. Overlap in *ti·* Vocabulary with Prose Vocabulary (Emeneau 1974b:7). 224

Figure 60. Code Switching in Tamil (Shanmugam Pillai) 275
Figure 61. Formal Aspects of the Terminology (Yalman 1962). 297
Figure 62. Terms of Address in Bengali (Das 1968) 316

Preface and Acknowledgements

The following work is an attempt to provide a modest overview of linguistic diversity in South Asia, and to place this diversity in a cultural context. The w[illegible]s largely bibliographic, and tries to describe the current state of knowledge concerning socially conditioned language variation in the subcontinent. The literature on such variation is large, and continually growing. Regrettably, it has been impossible to include much recent literature with this study. We have tried, however, to enumerate what we consider to be the major issues involved in a number of sorts of sociolinguistic diversity in South Asia, discuss the major literature written to date on those subjects, and state fruitful areas for future research.

Given the huge scope of this work it was inevitable that much of the material covered would be out of the academic specialization of either or both of the authors. A division of labor was, of course, necessary. Chapters 1, 2, and 4 as well as sections 3.0, 3.3, 3.4, and 7.2 were written by Michael C. Shapiro, whereas chapters 5 and 6, and sections 3.1, 3.2, 7.0, and 7.1 were written by Harold F. Schiffman.

Taking material from a large number of sources, we naturally encountered problems in abbreviations and transliterations. In general, we have not altered the transliterations given in direct quotations. We have, however, tried to standardize transliterations used in the body of the text. Abbreviations have generally not been changed in quotations, and should be clear from context. Abbreviations in the text, except where noted to the contrary, are noted on the following page.

ABBREVIATIONS

A	Apabhramsa
AV	Atharva Veda
B	Bengali
Bad	Badaga
Br	Brahui
Dr	Dravidian
G	Gujarati
Ga	Gadaba
Go	Gondi
H	Hindi
IA	Indo-Aryan
Ka	Kannada
Kod	Kodagu
Ku	Kumaoni
Kur	Kurukh
M	Marathi
Ma	Malayalam
Mai	Maithili
Malt	Malto
N	Nepali
Nk	Naiki
Oll	Ollari
Or	Oriya
P	Punjabi
Pa	Parji
Pas	Pashto
Pe	Pengo
Pkt	Prakrit
RV	Rig Veda
S	Sindhi
Sa	Santali
Si	Sinhalese
Skt	Sanskrit
Ta	Tamil
Te	Telugu
To	Toda
Tu	Tulu

There are a number of individuals whom we would like to thank for their contributions in producing this volume. Ms. Güliz Kuruoglu and Mr. Theodore Adams provided extensive assistance in much of the bibliographic research. We are grateful for their contributions. Ms. Leslie Adams helped in the early stages of preparing the manuscript. Special thanks go to Ms. Ruth Ross for enduring the thankless ordeal of typing and helping to edit the final manuscript. Her patience and good humor under duress helped make the completion of the manuscript possible. Final thanks go to Ms. Julia Petrov, Chief of the Research Program, International Studies Division, DHEW, for her encouragement in carrying out this project and for her indulgence in waiting for delivery of this long overdue project report.

Chapter 1

Introduction

South Asia is an area of staggering linguistic diversity. In its roughly 1,700,000 square miles (if one includes under the term "South Asia" India, Pakistan, Bangladesh, Sri Lanka, Nepal and Bhutan) are to be found languages and dialects spoken by roughly one-quarter of the world's population and representing at least five major language families and subfamilies--Indo-Aryan, Dravidian, Austro-Asiatic, Tibeto-Burman and Iranian. By general consensus each of these groupings encompasses numerous mutually unintelligible "independent languages," the names of which are in turn convenient labels for sets of mutually unintelligible and intelligible "dialects." This plethora of linguistic codes is confounded by an array of orthographic systems, themselves of diverse palaeographic origins. The widespread diffusion of these writing systems into different areas and their adoption by heterogeneous groups to represent different codes has led to definitional questions about what constitutes a language and how languages relate to orthographic systems.[1] Such confusion has complicated the description of language distribution in South Asia.[2]

In addition to being characterized by representatives of a multitude of linguistic stocks (a diversity which can be partially attributed to the complex ethnographic history of the region), the linguistic situation in South Asia is also sociologically complex. The religious, cultural, ethnographic, geographical, economic and political diversity of the area is well known and need not be spelled out here. Nevertheless it is fair to say that differences in many aspects of social structure are frequently reflected both in language use and in attitudes toward language use in South Asia. All

South Asian languages that have been recorded to date exhibit socially-determined structural differences.[3] All have myriad styles appropriate to particular contexts and in many cases limited in use to particular social groups.[4] South Asian languages in addition often contain more than one normative variety with the use of one or more of the alternatives confined to particular contexts. Literary Tamil, for instance, is based not on any current spoken form of the language, but rather appears to be derived from that variety of 13th century Tamil described in the grammar *nannūl*.[5] Spoken realizations of literary Tamil are not readily understood by those who control only the modern spoken dialect. The prevalance of such split norms, termed "diglossia" by Charles Ferguson (Ferguson, 1959), is so pronounced in South Asia as to undermine traditional attitudes about what constitutes a language and what a dialect, as well as to raise major pedagogical problems for bringing about literacy in South Asian languages.[6]

The linguistic diversity of South Asia extends even beyond these limits. Whenever an area contains a large number of mutually unintelligible language forms, techniques are required to enable different groups to communicate with one another. In such circumstances lingua francas (well-known examples of which include the Chinook Jargon of the coastal Indians of the Pacific Northwest of the United States, and the Swahili of East Africa), pidgins (roughly speaking, varieties of a language having a simplified grammatical and phonological system and which are not native to their users), and creoles (languages produced by the effective merger of two or more distinct codes and which become native to their users) are frequently developed and utilized. All of these have been used in South Asia.[7] It is also often found that languages not native to the area, and which consequently may be emotionally "neutral" to the speakers of some or many of the locally prevalent languages or dialects, are employed to facilitate intergroup communication. The widespread use of English among the educated in South Asia is an example of such a process.[8]

In addition to the development or utilization of such intergroup codes the linguistic diversity of South Asia is reflected in the existence of large numbers of bi- and multilingual individuals and groups. The linguistic competence of such individuals can be of many different varieties, ranging from that of one who is able to minimally understand sentences said to him in a second language but who is unable to orally produce any sentences in that second language, to another who commands all of the skills necessary for educated communication, both spoken and written, in several languages. Patterns of bi- and multilingualism represent one kind of adaptive reaction to a situation of extreme linguistic diversity. Any attempt to characterize the linguistic diversity in South Asia which actively seeks to describe the linguistic repertoires and habits of significant numbers of South Asians will have to come to grips with bi- and multilingualism. In fact, as we shall see in Chapter 6, multilingualism in South Asia is of a particular kind not usually described in the literature on multilingualism in the West.

As in many other parts of the world, linguistic diversity in South Asia has led to language-related political debates. Education must be carried on and literature, both popular and technical, must be published. National and regional governments, with the large bureaucracies they entail, must be administered. Communication networks, radio and television among them, must choose among available codes in their broadcasts.[9] The selection of one or more codes as the medium for these functions, at the expense of others, cannot fail but to have prejudicial implications for those who do not speak the codes selected.[10] Moreover, if such selections are carried out in different areas of social interaction, and if they repeatedly favor the same code or codes, then in the absence of countervailing social forces, there may be an alteration in the overall linguistic balance of an area; that is, one language may come to supercede another as, for example, standard French has come to replace Provençal and other dialects in France. One frequently finds social forces, often of unequal strength and effect, working for or against the use by groups of individuals of specific linguistic codes.[11] Social forces working in one direction may evoke counter-reactions, possibly resulting in confrontations characterized by animosity or even escalating into violence.[12]

It is clear then that definitional problems are immediately encountered in trying to rationally characterize the linguistic diversity of South Asia. The scope of the problem can be illustrated by trying to determine how many "languages" are spoken in South Asia, and how many "dialects". One immediately becomes enmeshed in such questions as: Are Hindi and Urdu one language or are they two?; Is Rajasthani a language or a dialect?; Is Ceylon Tamil a dialect of Tamil, or is it closer to Malayalam?; and, What do we mean by "Hindustani"? The common usages of the terms "language" and "dialect" are of little help in unravelling these problems. The term "language" has been used in many ways and often the linguistic entities designated by it are quite distinct (i.e. they may be paraphrasable as, among other things, literary language, spoken language, set of grammatical conventions, code recognized for official use by some geopolitical entity, code having significant literature written in it, code associated with institutionalized orthographic conventions, code used by some prestige group or groups of a society, etc.).

A quick examination of some of the entities popularly referred to as "languages" reveals how thorny the problem is. Chinese is commonly thought of as a language, and cited as the most widely spoken one in the world. Yet included under the cover of the term "Chinese" are a number of regional norms whose spoken forms are mutually unintelligible to one another.[13] Written Chinese, however, is based on none of these spoken varieties, and has its own unique grammatical structure. When one reads a Chinese text he is not reading Chinese, Mandarin, etc., but rather something in a distinct literary language, which is known to all educated Chinese, but quite apart from the living languages which they speak.[14] Norwegian and Swedish, on the other hand, are almost always thought of as distinct languages, in spite of the fact that many varieties of Swedish and Norwegian are mutually intelligible, and the fact that the two standard languages share a large portion of their lexicons and grammatical features.[15] In terms of any conceivable metric of linguistic relatedness, spoken Swedish and Norwegian are much closer to one another than are the so-called Chinese "dialects." Social and political questions of nationality have obviously played a role here in the popular determination of what is a language and what is a dialect.[16]

Similar problems are everywhere evident in South Asia. Prior to the independence of India and Pakistan in 1947, Punjabi was widely considered a dialectal form of Hindi, suitable for use in the home and its immediate environs, but passed over in favor of either Hindi or Urdu as a medium of literate exchange. With the exception of some groups of the Sikhs, vernacular education in the pre-partition Punjab was carried on either in Urdu or Hindi, with education in Sanskrit, Persian and Arabic prescribed for refining skills in either of these two languages. After partition, however, increasingly strong demands by many Sikhs for a Punjabi-speaking state led to the division in 1966 of Indian Punjab into a new, largely Punjabi-speaking state of Punjab and a Hindi-speaking state of Haryana. With this partition came the long-sought recognition of Punjabi as a medium of education in the schools, which in turn stimulated an increase in book publication in Punjabi (printed in the Punjabi, or *gurumukhī*, script) and in Punjabi radio broadcasting.[17] Whereas Punjabi before 1947 had, except for certain religious groups, been popularly thought of as a regional dialect of Hindi or Urdu,[18] political forces after partition led to a revision in popular attitudes toward the code which eventually resulted in political recognition of its status as a distinct "language." Other North Indian regional vernaculars, some of which had notable literature and which in some cases have independent scripts available to them, have never succeeded in gaining widespread recognition as independent languages. The case of Maithili, an Indo-Aryan "language" spoken in Bihar and areas of Southern Nepal, immediately comes to mind here.[19] It seems that the status of Punjabi as the liturgical language of Sikhism has given it an undeniable boost toward "language" status, an advantage that the other regional vernaculars lack.

The first step, then, in the sociolinguistic description of language use in South Asia is the establishment of terminology. The immediate problem is deciding what shall be considered a language and what a dialect. But a superficial examination of the scope of the problem will reveal that a two-way division of speech forms into languages and dialects is insufficient for sociolinguistic purposes. New terms must be provided, or consistent uses of old terms established, for correlating linguistically significant aspects of language use with social, political and geographical variables. What is one to call the linguistic competence of an individual whose repertoire of skills straddles what are usually considered distinct "languages?" Is there any sociolinguistic utility in constructing a term whose reference includes a heterogeneous set of linguistic skills and aptitudes? What are we to call those discrete aspects of an individual's total linguistic repertoire which are restricted to particular social contexts? Here such terms as "styles," "registers," and "speech varieties" come to mind, although general linguistic literature has been remarkably inconsistent in the use of these terms.

Similarly, what is one to call a common core of grammatical and phonological processes whose spoken outputs are conventionally associated with discrete systems of orthographic representation?[20] Should there be a term for referring to a body of phonological, morphological, or syntactic properties which is shared by a number of genealogically distinct speech varieties? In the particular case of South Asia, Bloch (1965), Emeneau (1956, 1969, 1974), Andronov (1964b) and others (cf. Chapter 4, "South Asia as a Linguistic Area") have pointed out a large inventory of linguistic features which are held in common by a majority of "languages" of the area, and which could not have independently emerged in each without the influence of neighboring speech forms. Are there terms available for referring to both the body of shared linguistic properties and the geographical area characterized by the dissemination of such linguistic features?

None of the problems enumerated above is readily answerable, and several immediately evoke questions of more general linguistic importance. The relationship of languages to dialects is, and traditionally has been, connected to the theoretical discussion of how languages change. Key questions in this matter have concerned the formal linguistic mechanisms of change, the diffusion of changes once initiated, and the resolution of structural tensions produced by the absorption of "non-native" elements into homogeneous systems. Of particular importance in these discussions has been the causality of sound change. In the nineteenth and early twentieth centuries an extensive literature was built up discussing whether linguistic innovations could be considered non-predictable spontaneous occurrences or whether they occurred as a result of what were tantamount to internal attempts to resolve structural tensions synchronically observable in grammatical systems.[21] Of equal importance to many linguists was the extent to which linguistic changes, once initiated, were invariable, and the degree to which they were susceptible to individual control.[22] Many nineteenth and twentieth century views on the oppositions between languages and dialects, as well as many of the specific typologies that were constructed for taxonomically classifying either the linguistic codes of a given area or the historically related members of a "language family," are natural outgrowths of *a priori* assumptions or stated theories on the nature of sound change. Moreover, most nineteenth century approaches to the dichotomy involved notions stemming from Romanticism, and many of them were based on the political and linguistic status quo of that time.

The definitional problems which are encountered in discussing linguistic diversity in South Asia are not narrowly linguistic ones. It is well known that many theories of general language typology use terms referring to various orders of social organization. Leonard Bloomfield, for instance, in his influential article, "A Set of Postulates for the Science of Language" conceived of language as "the totality of utterances that can be made in a speech community" and defined speech-community as a community "within [which] successive utterances are alike or partly alike. . ." (Bloomfield, 1926:26).

In other places it is not uncommon to find discussions of language and dialect phrased in terms of nationality and group identity. The general linguistic literature on social dialectology has made recourse to terms designating any of a number of orders of social grouping. All such associations of speech varieties with particular social orders have encountered difficulty, generally stemming from an inability to establish the referents for such sociological terms. Models developed by various linguistic schools of thought have managed to avoid specifying the nature of the social groups which use linguistic varieties by concentrating on the essential unity of the linguistic code itself.[23] To many linguists, linguistic analysis is made possible only by the assumption of a tightly interwoven object of description which is representable by a set of formalizable rules or principles. To such linguists, a speech community is constituted by the set of individuals who actively control this homogeneous set of rules, rather than being a preexistent entity which aids in the definition of speech forms.

As sociolinguistics is an intellectual enterprise which in part consists of the association of recurrent patterns of linguistic behavior with various sociological variables, it is important to have a clear and unambiguous means both for describing these variables and for characterizing the social groups among which they are distributed. It will not do simply to make use of assumed and undefined notions of "nation" or "people" and transfer them wholesale to the context of South Asian speech forms. The procedure is invalidated by the fact, among others, that many commonly used conceptions of nation and ethnic identity have only recently been born of the experiences of European and other Western states in development of their modern political structures. There is no necessary reason to believe that the emergence of modern political states in South Asia has or will have involved social groupings, alliances, and factions similar to those observed in the West.

As important as the development of a clear terminology for referring to social groups in South Asia is the establishment or determination of linguistically significant social variables. Western dialectology has made important strides in describing the geographical distribution of alternate speech forms. The study of the social distribution of speech forms, however, is in its infancy, and is only now in the process of determining appropriate variables for sociolinguistic investigation. The sorts of social variables which have been utilized in general sociolinguistic literature, i.e., age, education, economic status, ethnic background, type of employment, religion, I.Q., etc., are only a small subset of the variables which conceivably could be incorporated into the construction of meaningful generalizations about language use in South Asia.[24] Moreover, it is probable that the isolation of each of these variables raises general theoretical problems within sociological theory as well as special problems when they are adapted to South Asia.[25]

How, then, is one to begin a serious attempt to account for the linguistic diversity of South Asia? By way of an evasive answer, it could be suggested that an attempt be made to describe the patterns of language use of the area from the point of view of a trained linguist arriving in South Asia for the first time. Assuming that such a linguist does not possess prior knowledge of the results of Indo-European and Dravidian comparative and historical linguistics, and thus is not prejudiced by the weight of past linguistic classification, how would he be likely to describe the patterns of linguistic diffusion encountered in travelling across the region? Such a person would presumably be unprejudiced about which speech forms constitute languages and which, dialects. Although this book is largely an attempt to come to grips with issues of exactly this sort, it is nevertheless possible to reach some tentative conclusions about what this linguist might discover. First of all, for large portions of South Asia, most notably the great northern plain of India, the linguist would discover a virtual continuum of speech forms constituting a chain extending across the subcontinent from Karachi or Bombay in the west to Chittagong or Bhubaneswar in the east. On a "grass roots" level, adjacent links in the chain differ only by small-scale linguistic features. (The enumeration and description of what constitutes a "linguistic feature" is here left undefined although clearly it is theoretically undesirable to do so. Later in the book we shall attempt to describe the variables involved in differentiating among related speech forms). Such small-scale differences have little effect in influencing inter-group communication between adjacent links in the chain, but become compounded with greater distance. One is thus dealing with a diffusion of interconnected speech varieties among which mutual intelligibility is at least in some sense a function of distance. Among the dialects of this chain one can observe a great deal of structural similarity in phonology, morphology, syntax and vocabulary. Interspersed in this northern language continuum, our linguist would also be likely to find pockets of tribals (most notably in Bihar, Orissa, and West Bengal) whose speech forms seem linguistically unrelated to those of the surrounding districts.

In other parts of the subcontinent he would encounter large groups of dialects seemingly unrelated to the great group of speech forms of the northern plain. One such group effectively traverses the southern portion of the subcontinent and includes a multiplicity of language varieties showing significant coherence in phonology, morphology, syntax, and vocabulary. In many of the peripheral areas of the subcontinent he might observe groupings of speech forms which are less homogeneous in overall typology than observed in the larger language groupings to the south, and which clearly represent the juxtaposing of speech forms of a number of distinct origins. Within each of the major "blocks" of speech forms which one encounters in South Asia, the degree of linguistic differentiation is clearly less than is generally present between linguistic varieties cutting across these groupings. This last claim is not an unqualified one, however, as blurred transitional zones between the major linguistic groups often show patterns of bi- and multilingualism, wholesale lexical, and in some cases phonological and morphological, borrowing between codes, and the use of compromise codes to effect intergroup communication.

Such a characterization of language use in South Asia would only scratch the surface. The first observations given above are only a crude description of one level of language use, something restricted to a substratum of Indian society. They would relate, and highly inaccurately, the language use of a largely non-urban, uneducated element of society, whose social contacts are largely restricted to an area within a small distance from their birthplaces. The language use so described would then tend to be unaffected by the standardizing tendencies of the mass media and government educational policies. In the South Asian context our linguist would quickly discover that several layers of superstructure are superimposed over this basic network of speech forms. It is also clear that the structural patterns of the upper strata of language use differ radically from area to area within South Asia. In north India for instance, our linguist would find out that the local speech varieties seem to group themselves into a number of subregional norms (Bundeli, Avadhi, Bhojpuri, Maghi, Braj, etc.) with distinct literatures being composed in a number of these varieties.[26] These subregional varieties themselves are grouped by their speakers and by political institutions into larger-order regional "dialects" or "languages" such as Punjabi, Hindi, Gujarati, "Bihari," etc., but with widespread disagreement and confusion as to the proper groupings of sub-regional "dialects" into "regional languages."[27] This confusion is only exacerbated by governmental policies which integrate questions of linguistic classification with questions of state organization and reorganization.

In the southern ("Dravidian") portion of the subcontinent, on the other hand, there seems to be a much clearer vertical arrangement between the substructure and an identifiable level of "languages" under which the spoken varieties are subsumed.[28] The multileveled regional set of classificatory conventions employed in the north is greatly simplified in the south. In its place, however, we encounter a more extensive network of socially conditioned speech alternates, grouped under the rubric of single languages, than is obvious in the north. Thus under the cover of such terms as "Tamil" and "Telugu" are linguistically distinct speech forms conditioned by caste and other social variables. The existence of social conditioning factors for linguistic variation is widespread within the subcontinent, but with the particular social variables having different force in different regions.[29]

Pursuing his investigation, our linguist would quickly learn that once the factors of literacy and literary traditions are entered into the picture, the construction of multidimensional models of language use in South Asia is further complicated. In north India, for instance, the problem is particularly troublesome. Often speakers, who on the level of the substratum speak roughly comparable speech forms, will receive their formal education through what amounts to different educational systems and consequently will be instilled with different normative values. A simple example here would be speakers of roughly comparable colloquial vernaculars (i.e. some form of "Hindustani"), where one gets his education from the medium of formal "Hindi" and the other, "Urdu." In the former case, literacy is introduced through the *devanāgarī* script, a left-to right derivative of Brahmi, the script used in the Ashokan rock edicts, and the individual's technical vocabulary is supplemented with a large number of items directly borrowed from Sanskrit *(tatsamas)*. In the latter case, literacy is achieved through the use of the Perso-Arabic script, a right-to-left orthographic system which is radically different in its structure from *devanāgarī*, and the inherited vocabulary is supplemented with Persian and Arabic loanwords. In the case of these two individuals, the literature and printed materials studied in school draw upon two distinct literary and cultural traditions, from the Sanskritized "*shuddh hindī*" tradition on the one hand, and from the Perso-Arabicized "*sāf*" or "pure") Urdu tradition on the other. Education will here have produced polarization where on the level of our substratum there was relative unity.[30]

In South India, on the other hand, literacy and literary traditions affect linguistic diversity in ways fundamentally different from the ways they do in the north. For many south Indian languages there exists a deep cleavage between the structure of the literary language and that of the vernaculars. This cleavage in some cases is so great as to require that the literary language be learned virtually as a second language requiring considerable school instruction.[31] In many cases spoken realizations of literary language are unintelligible to speakers of non-prestige forms of the vernacular. Such stylistic cleavages, often resulting from attempts to enhance the sanctity or purity of some linguistic code by infusing it with linguistic elements from an external source (i.e. Sanskrit, English, Persian), or by purging from it all "non-native" elements and borrowing copiously from earlier lexical and grammatical forms of the language, adds an extra dimension to models postulated for describing the linguistic repertoires of groups of South Asians.[32]

It should be obvious from this discussion that the sociolinguistic investigation of South Asia presupposes a number of techniques. First of all, it requires a method for measuring degrees of linguistic relatedness. What does it mean to state that two languages are related to one another, and how can we measure degrees of relatedness? How do the various kinds of linguistic diversity encountered in South Asia, i.e., that among sub-regional varieties within the "Indo-Aryan" group, among the socially conditioned normative varieties of Tamil, between Sanskritized Hindi and Arabicized Urdu, etc., fall on such a scale? Moreover, how can scales be constructed which will serve for the comparison of linguistic repertoires substantially more heterogeneous than the codes traditionally dealt with in grammatical analysis? Additionally, such a sociolinguistic investigation of South Asia requires formal linguistic conventions and theories for providing synchronically adequate descriptions both of the "linguistic" features characterizing the repertoire of the individual and for characterizing the shared linguistic features of sociologically meaningful groups.

There are a number of important reasons for stressing the methodological points just stated. South Asian languages and dialects have been the subject of a vast amount of typological literature (cf. Chapter 3, "Traditional Taxonomies of South Asian Languages"). Genetic models of linguistic history have been postulated and modified numerous times during the past three centuries for each of the major typological linguistic stocks in the area. The results of these investigations have been impressive. For the Dravidian and Munda language groups, we have extensive and detailed Stammbaum family histories in which successive stages of the family can be related to one another by systematic phonological and morphological processes. In the case of Dravidian we have independent textual evidence to aid and facilitate the validation of historical reconstructions. For Tibeto-Burman there has been a substantial amount of work done in establishing family tree hierarchies and correlating differences in these tree structures with linguistic variables. In the case of Indo-Aryan and Iranian, Stammbaum diagrams have been postulated and revised, with an impressive array of data demonstrated to relate successive stages of the various language families and sub families. Extensive lexical correspondences have been pointed out between the major "languages" (cf. Turner, 1966) and the phonological and morphological changes in the development of particular modern standard vernaculars detailed (cf. Bloch, 1914 for Marathi; Chatterji, 1926 for Bengali, etc.).

These significant contributions, however, have been arrived at through great cost. One of the most obvious weaknesses has been either a failure to seriously specify the criteria used to differentiate between languages and dialects, or the adoption of criteria which are in one or more ways inadequate for dealing with broad questions of linguistic typology. It is not unfair to say that much of the Stammbaum tree reconstruction which has been carried out in South Asian linguistics has either presupposed the existence of homogeneous codes for description or paid lip service to the complexity of linguistic codes while in practice acting as if the codes were homogeneous. This fact has led to the inability of South Asian linguistics to talk adequately about the mutual influence of codes upon each other when placed in close contact. For sociolinguistic purposes we need to know in what ways these are modified in close juxtaposition, and what the mechanisms are by which these modifications take place. From the point of view of the single speaker, we need to know in what ways the presence of multiple codes, or single codes having multiple levels influences the individual's total linguistic competence, and to what ends and in what contexts the individual uses particular components of complex linguistic repertoires.

There are numerous other questions which have been slighted by the adoption of an uncritically Stammbaum approach to linguistic diversity in South Asia. What formal devices can be used to capture the societally conditioned aspects of language use? How do the total speech repertoires used by individuals relate to the structures of hypothetical homogeneous codes postulated by linguists and presupposed in determining what constitutes a standardized language? In what ways can an understanding of linguistic repertoires, when viewed in their social contexts, help us to define more carefully what we mean by such terms as "language" and "dialect?" And how can the description of language in its social context enable us to characterize the linguistic diffusion of a geographical area such as South Asia better?

The present volume to a great extent attempts to deal with the sorts of questions just enumerated. We are here interested in finding viable techniques for describing the sociolinguistic diversity of South Asia. We will discuss and evaluate the kinds of generalizations which a linguist can come up with if he does not presuppose the homogeneity of linguistic codes and the pre-existence of hierarchically arranged classificatory networks of "languages" and "dialects." The focus will be on precisely those aspects of linguistic diffusion in South Asia where the borders between these entities are haziest and the commonly used definitions most vunerable to close scrutiny.

"The Theoretical Description of Language and Dialect" is an attempt to enumerate the criteria which have been used in the past in the taxonomy of language forms, discussing a number of theories of dialect formation both from the point of view of linguistic innovation and that of diffusion of language change. We discuss the terminological frameworks for categorizing speech varieties entailed by different views of language change. The claims, conclusions, and assumptions involved in each of these sets of views are examined vis-à-vis the linguistic situation in South Asia. In addition, possible future applications of recent advances in sociolinguistic theory to the taxonomy of South Asian language varieties are proposed.

Chapter 3, "Traditional Taxonomies of South Asian Languages," is a rapid overview of past literature on the classification of South Asian languages.

We consider the theoretical frameworks within which these classifications were carried out, and point out areas of contention about the proper placement of certain languages and groups of languages. Also, an assessment is made of the limitations and weaknesses of much traditional classificatory work on South Asian languages.

In Chapter 4, "South Asia as a Linguistic Area," we turn our attention to the consideration of South Asia as a distinct linguistic area (or Sprachbund). This includes discussion of a number of sets of criteria which can be set up in establishing "linguistic areas" and evaluate the data on South Asia in the light of these criteria. In addition to analyzing the data which are used to support claims of the existence of a general South Asian Sprachbund, we attempt to evaluate evidence pointing to the possibility of creating one or more "micro-linguistic areas." The discussion in this chapter focuses on general typological theory and proceeds to an evaluation of the specific phonological, morphological and syntactic bases for establishing one or more South Asian linguistic areas.

In Chapter 5, "Social Dialectology," we evaluate the literature on South Asian social dialects in view of the theoretical and terminological discussion of Chapters 3 and 4. After summarizing the literature which can substantiate the association of linguistic variables with social ones, we examine the extent to which such associations are possible in South Asia, and in so doing consider the possible applications of general sociolinguistic theory to the specific problems of the subcontinent.

Chapter 6, "Individual and Group Linguistic Repertoires," is an examination of linguistic codes which encompass elements from what traditionally have been considered more than one autonomous language. The chapter includes a general theoretical discussion of the literature on bi- and multilingualism. We describe the state of current knowledge about bi- and multilingualism in South Asia, concentrating on the linguistic processes which are demonstrated in the phenomena. This is followed by a discussion of the literature on pidgins and creoles in South Asia. Particular emphasis is placed on pidginization and creolization as highly productive processes, or sets of processes, arising in situations of language contact. We consider the importance of pidginization and creolization, both for the development of "linguistic areas" and as integral parts of language change, and devote some attention to the roles which pidginization and creolization have played in the emergence of modern patterns of South Asian language and dialect diffusion. This chapter draws heavily from the theoretical literature on pidgins and creoles in other parts of the world.

The subject matter of Chapters 2-5 essentially concerns relationships holding between autonomous codes. Chapter 4, for instance, deals with the possibility of abstracting sets of linguistic features from sets of geographically contiguous codes, and of establishing geo-linguistic areas as functions of the sharing of these features. Chapter 5 largely concerns the association of linguistic variables, prior to the establishment of fixed entities to be referred to as "languages" or "dialects," with definable social variables. The material on pidgins and creoles in Chapter 6 deals with what amounts to implicit attempts to construct intercode systems to facilitate interaction between groups commanding alternative modes of communication.

In the first part of Chapter 7, "Ethnographic Semantics and the Ethnography of Speaking," our attention is turned to the ways in which the lexicons of South Asian "languages" and "dialects" contain elements which structure themselves into concrete systems. Traditionally such studies, commonly focusing on so-called "kinship" systems, have been included in the domain of social anthropology. Anthropologists have tended to view the lexicon as a possible means of approach to major generalizations about a society or culture. They have assumed that structural aspects of a society may be reflected in subsystems of the language or languages spoken by the members of that society or culture.[33] Anthropologists have commonly generalized from the existence of such structural subsystems to the existence of related properties of society. In this book, however, our interest is not in the verification of hypotheses about society through linguistic patterning, but rather in the determination of the extent to which such lexical patterning is present, and also in the extent to which its presence (if such presence can be demonstrated) aids in the demonstration of "linguistic areas" for South Asia, as well as sheds light into our understanding of linguistic diffusion of the area. Our aim, then, is to discuss the linguistic aspects of "ethnographic semantic" systems only as they pertain to more general linguistic questions and not as a strategy leading toward a general sociological description of South Asia.

The latter portion of Chapter 7 deals with the functional purposes for which a single code can be employed. Admittedly, this is an area in which there has been very little research to date, although it is an area rich in potential. A recently completed doctoral dissertation by Dhanesh Jain (Jain, 1973), in part dealing with the socially conditioned uses of morphologically distinct elements of the Hindi pronominal system, is a preliminary attempt to deal with such questions of linguistic pragmatics. Examples of topics discussed by Jain are the ways in which a language uses circumlocution and/or systematic evasion to avoid directly specifying the names of individuals standing in particular social relations to the speaker. Jain convincingly demonstrates that Hindi contains a tightly structured system with regard to the naming or lack of naming (termed "no-naming") of individuals which builds upon the morphologically marked pronominal system of the language. This discussion is clearly related to the discussion of honorifics as a semantic/morphological category of a language, but is substantially broader than it. In addition to Jain's works, we also enumerate several other studies which have data pertinent to our understanding of the ethnography of speaking in South Asia.

NOTES: CHAPTER 1

1. The case of Hindi and Urdu is the most notorious problem. Many individuals consider Hindi and Urdu to be variants of the same language, with the former written in the *devanāgarī* alphabet and the latter in a modified form of Perso-Arabic script. To such individuals the differences between the two speech forms are stylistic, and consist primarily of overlapping stocks of lexical items and slightly different phonemic and morphological inventories. It is equally possible to consider them two distinct languages, and to support this view by pointing to the considerable differences in their literary registers as well as by enumerating ever-widening discrepancies between the official versions of the codes when used by the respective governments of India and Pakistan. The role of script in creating this confusion is illustrated by the fact that Premchand's novels, which are variously printed in either *devanāgarī* or Perso-Arabic script, are cited as classics of both Hindi and Urdu literature. This is in spite of the fact that the language of his novels is typically a colloquial "Hindustani" readily understandable to most Hindi and Urdu speakers, regardless of how they identify their mother tongue.

2. This is particularly so in determining which speech varieties are merely "dialectal" variants of others. If it can be shown that a speech variety is associated with a distinct script, it is easier to assert that that variety constitutes an independent language, and, consequently, that its speakers should be accorded certain rights and benefits. (Cf. p. 4 for a discussion of this type of situation in the Punjab).

3. For instance, virtually all have multiple second person pronominal forms, with the selection of forms for particular address and referential functions dependent upon the absolute and relative social positions of the speaker and addressee, the state of mind of the speaker, and the amount of respect which the speaker wishes to accord the addressee. For full discussion of the socially conditioned use of pronominal forms see section 7.2.2.

4. E.g. B. *sādhu bhāṣā* "literary language" and *calit bhāṣā* "colloquial language" (cf. Dimock, 1960); literary and colloquial Tamil (cf. Shanmugam Pillai, 1960); and Brahmin and non-Brahmin Kannada (cf. Bright, 1960b).

5. Cf. Venkatarajulu Chettiyar, 1959:319-20; and Zvelebil, 1964:250.

6. If literary Tamil is substantially different from the colloquial, then children educated in the former are being asked to learn the equivalent of a second language. This means that instead of merely learning to associate a set of graphic symbols with linguistic units already known to them, they are faced with the vastly more difficult problem of learning a new set of linguistic conventions as well. This is essentially the same problem as occurs in education in modern Arabic, where spoken dialects differ considerably, but where all share a common literary language which has been artificially preserved for centuries. Arabic children must all learn an archaic form, only obliquely related to their spoken dialects, in order to become literate (cf. Ferguson, 1959).

7. Cf. 6.4.

8. Cf. 6.5.

9. This process of selection can often involve heated political controversy, as the use of a code in certain types of communication can become a powerful symbol for achieving group recognition. For instance, in the unsuccessful drive for political recognition of Maithili, one of the earliest and most persistent demands was for the creation of a Maithili language radio station at Darbhanga, Bihar (cf. Brass, 1974:112-14).

10. For instance, this is the case in the administration of entrance examinations for various types of government service. It has been argued that if it is possible to take entrance examinations for the Indian Civil Service in Hindi rather than in English, then the speakers of languages other than Hindi are being unfairly discriminated against, since neither Hindi nor English is their native tongue. The retention of English for these examinations, however, is strongly criticized

because of its colonial associations, and because such a practice tends to leave control of the government agencies in the hands of the small minority of individuals who are already proficient in the language (cf. Das Gupta, 1969; Friedrich, 1962).

11. In much of North India, for instance, forces ranging from the government educational policy and All India radio broadcasting to nationalist sentiments arising after political separation from Pakistan have all led to increased Sanskritization in Hindi, and to a reduction in pejorative connotations of the use of the code by males outside of their homes. On the other hand, the wide dissemination of Bombay "Hindi" movies, which are highly Persianized in language and thematic content, provides a powerful countervailing force for the retention of Perso-Arabic elements in North Indian vernacular languages.

12. For example, when the government of Ceylon passed the Official Language Act in 1957, declaring Sinhalese to be the sole official language of the country, communal riots broke out between Tamils and Sinhalese. As the situation deteriorated, the government agreed to grant official recognition to Tamil as "the language of a national minority of Ceylon" (Kearney, 1967: Appendix II). This agreement prompted demonstrations led by Buddhist *bhikkus* who felt that the cause of Sinhalese had been betrayed. Resulting pressure from the Sinhalese majority led to abrogation of the agreement concerning recognition of Tamil, and aroused communal sentiments again resulting in widespread rioting and violence (cf. Vittachi, 1958).

13. Cf. Bloomfield, 1933:69.

14. Cf. Martin, 1972.

15. Cf. Haugen, 1966.

16. For an excellent discussion of how notions of nationality have influenced definitions of language and dialect, see Haugen, 1966a.

17. Cf. Brass, 1974:277-400; Das Gupta, 1970:152-7.

18. Prior to 1947, Punjabi was commonly thought of as a "house language," suitable for use in the home and its immediate environs, but was not considered an entity to which one owed conscious allegiance or which one used for educational or business purposes. This attitude was widely reflected in census returns, which commonly indicate the codes which individuals report as being their native tongues, or which are standardized by census takers and supervisors. Under such circumstances there was little formal pressure to record Punjabi as a native tongue. (The autonomy of Punjabi from *khaRĪ bolĪ* Hindi was, however, recognized by professional linguists (cf. Brass, 1974:286-91)).

19. Cf. Brass, 1974:51-116.

20. As, for example, various styles of colloquial "Hindustani" which can be written equally well in *devanāgarĪ* or Perso-Arabic script.

21. Cf. Pedersen, 1931; Robins, 1967:164-97.

22. Cf. 2.2.

23. For a lucid summary and critique of these views, see Weinreich, Labov, and Herzog, 1968.

24. Unfortunately the variables which have been applied are largely those which have proven fruitful in western social dialectology, or which seem obvious from the anthropological and sociological study of South Asia. The determination and validation of productive variables for sociolinguistic purposes is a major desideratum for future research.

25. It is necessary that these variables be independently motivated, and that their use have predictive power for other non-linguistic aspects of South Asian social structure.

26. Cf. Cardona, 1974; Mishra, 1971; Tiwari, 1969:185-310.

27. Cf. 3.1.

28. Cf. 3.2.

29. Cf. Chapter 5.

30. Cf. 5.1.

31. "A Kannada speaker requires about seven years of formal schooling before he can understand the literary dialect completely, and some practice or additional years of formal education may allow him to master it as a spoken style" (McCormack, 1960:80).

32. For discussion of the attempt to "purify" Tamil, see Schiffman, 1973a and Shanmugam Pillai, 1960.

33. A clear statement of this doctrine is in Emeneau, 1949.

Chapter 2

The Theoretical Description of Language and Dialect

2.0. Introduction

As it is impossible to talk about language in South Asia without considering the criteria used to establish what is a "language" and what a "dialect," we find it necessary to present here a short excursis on various models which have been used in the past to differentiate among languages, dialects, and other species of linguistic codes. We examine the criteria which have been used both to differentiate among codes and to group together related codes. We further discuss a number of models of linguistic change which have been postulated for describing the development of dialect differentiation, and analyze theories of language and dialect both from the point of view of synchronic areal study and from an historical perspective. As the goal is to provide a theoretical framework within which to explore the linguistic diversity of South Asia, we wish to describe and evaluate as wide a variety as possible of terminological conventions which have been used to categorize various orders of linguistic systems in the area. Such explication is necessary to enable us to construct meaningful critiques of specific analyses aimed at various aspects of South Asian language use.

Given the size and limitations of this volume, it is not possible to discuss all theories proposed for explaining the processes of dialect differentiation,[1] and an attempt has been made to choose among the most cogently argued and influential of these theories, particularly those which have been employed in the language classification of South Asia. In this chapter our discussion is limited to five major theories (or sets of theories), most of which concern both language change and the linguistic diversity resulting from it. In a number of cases these theories discuss the notion of linguistic *relatedness* and seek to specify devices and procedures for measuring relatedness among speech varieties. In section 2.6 we try to point out some of the more obvious strengths and weaknesses of each of these approaches for describing language use in South Asia, identifying aspects of South Asian language use which clearly fall outside the descriptive potential of all of the discussed frameworks for handling linguistic diversity, and briefly indicating the extent to which each of these approaches has played a role in forming prevalent attitudes about what constitutes a language and what a dialect in South Asia.

2.1. Isoglossal Theories of Language and Dialect

During the first half of the twentieth century a number of European scholars, both drawing upon and reacting to the huge body of literature dealing with comparative Indo-European linguistics and dialectology,[2] sought to develop a framework within which they might discuss linguistic differences between related speech forms as well as assign labels to various orders of linguistic codes. A basic assumption of all such schools was that it is possible to isolate "minimal" linguistic features which may correspond to any overtly marked aspect of linguistic structure--phonology, morphology, syntax, etc.--and which may be used to differentiate two non-identical speech varieties. The occurrence of any of a number of alternate values of a linguistic variable can be graphically located on a map of the geographical area in which the alternate speech forms are used. Repeating this procedure for each of a finite number of geographical points thus produces a network of values with regard to the significant variable.

A sample of such a geographical network in New England is pictured in Figure 1. Each instance of the marks +, ∘, or • indicates the occurrence of a regional version of the lexical item *cottage cheese*, whose alternate forms are *pot cheese, sour milk cheese* **and** *Dutch cheese*. Once such a graphic grid has been constructed it is possible to represent the pattern of distribution of alternate values of these features by drawing lines on a map in such a way that they separate geographical areas which differ with regard to the identified linguistic features. These lines, termed "isoglosses," roughly speaking indicate the

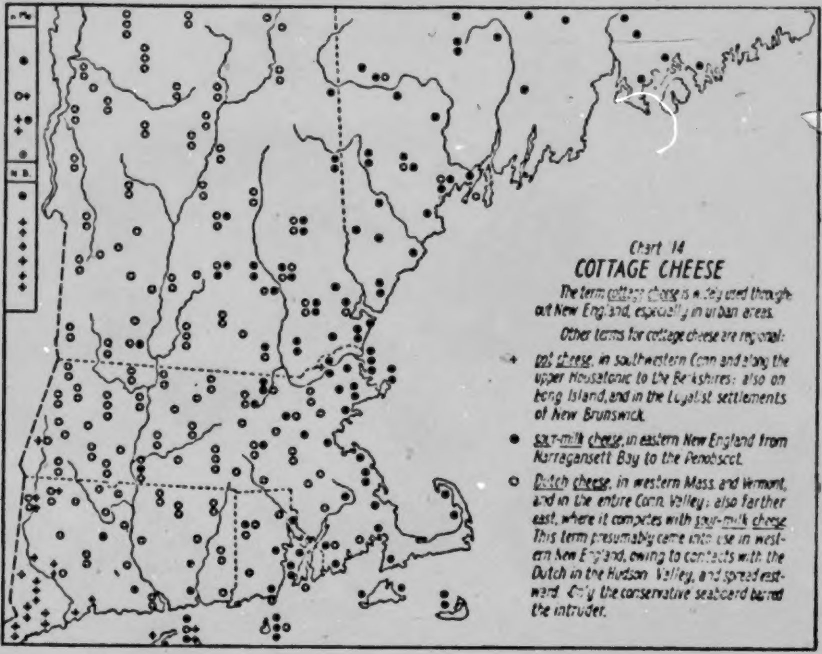


Chart 14
COTTAGE CHEESE
The term cottage cheese is widely used throughout New England, especially in urban areas.
Other terms for cottage cheese are regional:
+ pot cheese, in southwestern Conn and along the upper Housatonic to the Berkshires; also on Long Island, and in the Loyalist settlements of New Brunswick.
● sour-milk cheese, in eastern New England from Narragansett Bay to the Penobscot.
○ Dutch cheese, in western Mass. and Vermont, and in the entire Conn. Valley; also farther east, where it competes with sour-milk cheese. This term presumably came into use in western New England, owing to contacts with the Dutch in the Hudson Valley, and spread eastward. Only the conservative seaboard barred the intruder.

Figure 1. Alternate forms for *cottage cheese*
(From Kurath, 1939:33).

geographical limits of the dissemination of linguistic features, and, by extension, also specify the limits of a speech community or dialect area. The drawing of such lines allows the establishment of what can be considered "dialects" *with regard to the specific feature dealt with by the isogloss*.

Once isoglosses have been drawn for a large number of linguistic features, observations can be made about their pattern of congruence. Areas separated from one another by only a single isogloss are considered to have greater affinity than areas divided by a convergent "bundle of isoglosses." It is then possible to consider the degree of relatedness of geographically contiguous speech forms to be inversely proportional to the density of the bundles of isoglosses separating them.[3] A ready example of these phenomena is provided by data on German village dialects. The map in Figure 2 is considered to represent a number of Swabian villages centering around Bubsheim. The ten lines or isoglosses indicate the dissemination of sets of alternate phonetic realizations of ten lexical items. The point of reference of the chart

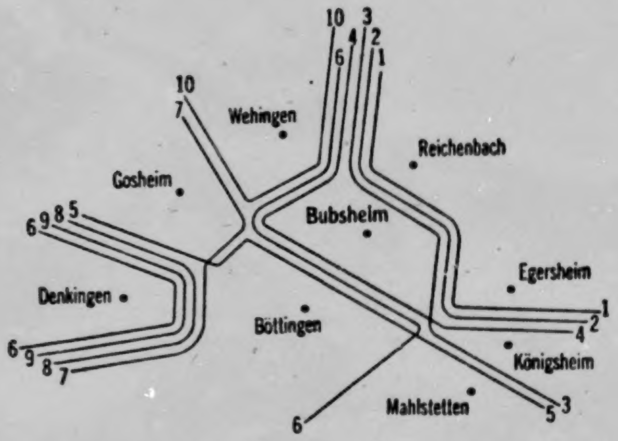


Figure 2. Isoglosses in ten Swabian villages (taken from Haag and cited in Bloomfield, 1933: 326).

is that of the village of Bubsheim, with the pronunciation of surrounding villages related to it. (A synopsis of the phonetic differences charted in Figure 2 is given in Figure 3.)

The use of isoglosses in this manner facilitates the formulation of a number of generalizations about the relatedness of the village speech forms. Reichenbach and Egersheim, as a unity, are maximally differentiated from Wehingen, being separated by six isoglosses, while Böttingen and Mahlstetten, separated by only one isogloss, have virtually homogeneous speech forms.

The isoglosses given in Figure 2 are derived by comparing alternate phonetic realizations of discrete lexical items. In theory such isoglosses can be drawn to demarcate any linguistic features (lexical, morphological, syntactic, phonetic, etc.) by which two speech forms can vary. In practice, however, rather than attempt the staggering task of plotting isoglosses to mark the distribution of the alternates of every point of variation in a language,[4] many dialect geographers attempt the more feasible task of plotting the distribution of alternates of second order variables. (An example of this is plotting the occurrence of *b* versus *p* in initial position in words rather than plotting numerous graphs of the distribution of *p* and *b* in particular words starting with either of these sounds.) Thus Robert A. Hall, in attempting to describe the dialect situation of Franco-Provençal, plots

Bubsheim	Reichenbach, Egesheim	Königsheim	Mahlstetten	Böttingen	Denkingen	Gosheim	Wehingen
1. ofə 'stove'	o:fə						
2. uffi 'up'	nuff						
3. tsi:t 'time'	tsejt	tsejt					
4. baw^{n} 'bean'		bɔ:^{n}	bɔ:^{n}	bɔ:^{n}	bɔ:^{n}	bɔ:^{n}	bɔ:^{n}
5. ɛ:^{n}t 'end'			aj^{n}t	aj^{n}t	aj^{n}t		
6. me:jə 'to mow'				majə		majə	majə
7. farb 'color'					fa:rb	fa:rb	
8. alt 'old'					a:lt		
9. truŋkə 'drunk'					tru:^{n}kə		
10. gaw^{n} 'to go'							gɔ:^{n}

Figure 3. Phonetic differences in the pronunciation of ten German lexical items (from Bloomfield, 1933:327).

five phonological alternations in terms of isoglosses, where each isogloss has presumably been generalized from a large number of "single feature" isoglosses (see Figure 4).

The notions of isogloss and bundling of isoglosses have traditionally been used by dialect geographers to give substance to the dichotomy of language and dialect. It has been assumed that geographical areas show hierarchical patterns in the convergence of isoglosses.[5] The main dialect boundaries of an area are simply those having the greatest concentration of isoglosses. On the basis of the density of bundles of isoglosses it is possible to group regional speech varieties into higher order regional dialects, which can in turn be grouped into "languages." Samples of the establishment of regional dialect areas as a function of the thickness of bundles of isoglosses are shown in Figures 5 and 6.

In a number of versions of isoglossal theory it is either implied or explicity claimed that thick bundles of isoglosses most often correspond to geographical or political boundaries.[6] This is itself derivative of a view which holds that linguistic diversity is a function

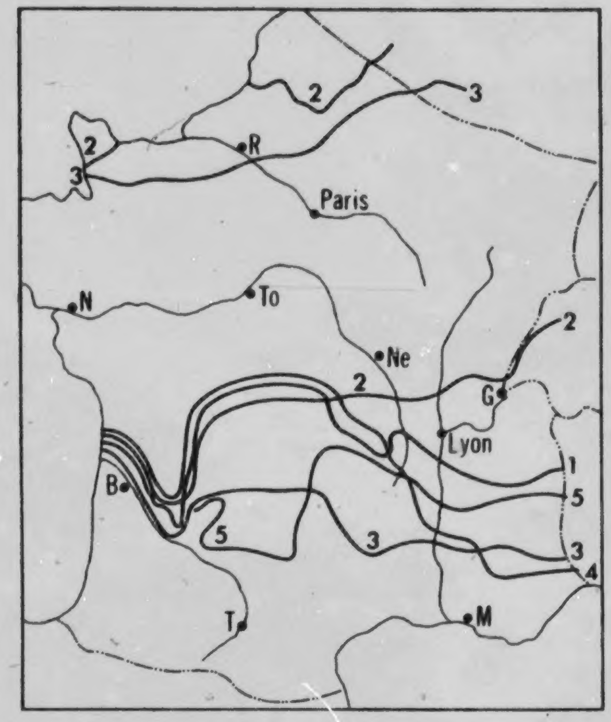


Figure 4. France: Outer limits of northern phonological features.

1. Northern *e* (as in *cher*) ≠ Southern *a* (< L. *cārus*)
2. Northern *e* (as in *mère*) ≠ *a* (< L. *māter*)
3. Northern *ča-* (> *š*, as in *champ*) ≠ Southern *ka-* (< L. *campus)*
4. Northern loss of *-d-* (as in *chaine*) ≠ Southern *-d-* (< L. *catēna*)
5. Northern loss of preconsonantal *s* (as in *château*) ≠ Southern preservation of *s* (< L. *castellum*)

Cities: B(ordeaux), G(eneva), L(yon), M(arseille), N(antes), Ne(vers), P(aris), R(ouen), T(oulouse), To(urs)

(From R. A. Hall, Jr. "The Linguistic Position of Franco-Provençal," cited in Kurath, 1967:97).

of the lack of social or group interaction.[7] The greater the social interaction, it is claimed, the greater the centripetal force toward linguistic uniformity. Natural boundaries such as mountain ranges and rivers and man-made political boundaries serve to isolate social groups, thereby leading to linguistic diversity. This diversity is characterizable by thick bundles of isoglosses overlaid on a geographical map of the area.[8]

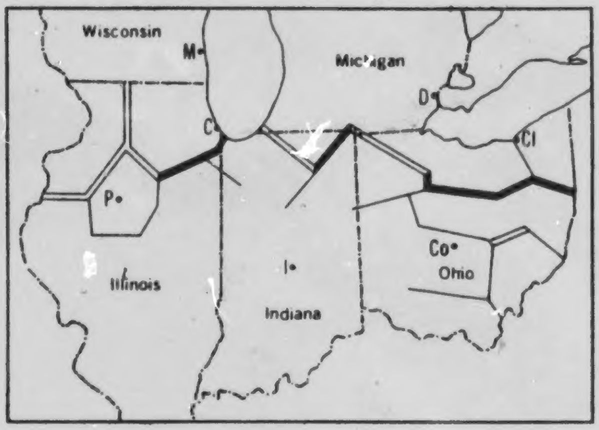


Figure 5. Bundles of heterolexes in the Great Lakes Area.

━━━ more than 15 heterolexes
═══ 10-14 heterolexes
─── 5-9 heterolexes

Cities: C(hicago), Cl(eveland), Co(lumbus), D(etroit), I(ndianapolis), M(ilwaukee), P(eoria).

Adapted from A. L. Davis. "A Word Atlas of the Great Lakes Region," and cited in Kurath, 1967:30.

In terms of this model of dialect geography, distinct languages are speech forms characterized by maximal differentiation and indicated by maximally thick bundles of isoglosses separating the areas in which they are spoken. The model implies that it is possible to create a linear scale of relatedness among speech forms ranging from total congruity (i.e. not separated by any isoglosses) to non-relatedness (as characterized by thick bundles of isoglosses).[9] Languages and dialects are considered to refer not to specific parts of the scale, but to relative positioning on it.

There are a number of general observations which need to be made about this model of language and dialect. The most important is that its apparent validity diminishes rapidly under careful scrutiny. Virtually all of its applications have been to cases where there has never been any serious debate about whether the speech forms being analyzed are languages or dialects. In the cases of European and American dialectology, the problem has never been to ascertain how many distinct languages are spoken in a given area, but rather to delimit the

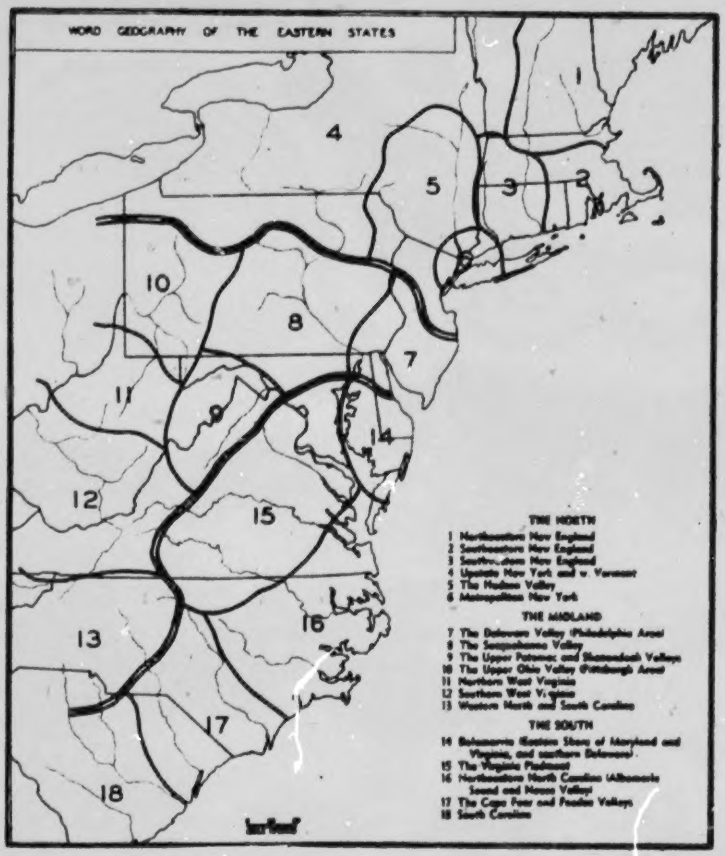

WORD GEOGRAPHY OF THE EASTERN STATES
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
THE NORTH
1 Northeastern New England
2 Southeastern New England
3 Southwestern New England
4 Upstate New York and w. Vermont
5 The Hudson Valley
6 Metropolitan New York
THE MIDLAND
7 The Delaware Valley (Philadelphia Area)
8 The Susquehanna Valley
9 The Upper Potomac and Shenandoah Valleys
10 The Upper Ohio Valley (Pittsburgh Area)
11 Northern West Virginia
12 Southern West Virginia
13 Western North and South Carolina
THE SOUTH
14 Delmarva (Eastern Shore of Maryland and Virginia, and southern Delaware)
15 The Virginia Piedmont
16 Northeastern North Carolina (Albemarle Sound and Neuse Valley)
17 The Cape Fear and Peedee Valleys
18 South Carolina

Figure 6. Speech areas of the eastern states.

Adapted from Hans Kurath, *Word Geography of the Eastern United States*, and cited in Kurath, 1967:28.

range and nature of regional dialects of a language whose existence is presupposed.[10] The model has, therefore, concentrated on sorting out relations among speech varieties along the end of the scale marking the greatest degrees of linguistic relatedness. The model implies a position about determining language boundaries, but has virtually never been put to the test of determining these boundaries where they were not already known. That such applications have not taken place is not accidental. The features by which non-related languages differ, as well as those which are shared by related language varieties are of numerous sorts. A theory which in essence reduces all order of linguistic features to a uniform set, so that they can be represented by isoglosses, has no basis for meaningfully characterizing the *structural* relatedness or lack of it among systems.[11] Moreover, the model has little means for qualitatively sorting out the linguistic features by which speech forms vary; it needs an external set of evaluative criteria before judgments can be made as to the relatedness of speech forms. Such criteria have not been forthcoming with the field of dialect geography, and it is therefore not surprising that the application of the model has taken place in cases only where it is possible to presuppose the relatedness of codes.

2.2. Stammbaum and Wave Models of Linguistic History

As is well known, one of the major advances of nineteenth and early twentieth century linguistics was the determination of familial relationships among the members of the so-called Indo-European family of languages.[12] The work done in this tradition was designed to establish hierarchical relations among the languages of this family and to relate the chronologically later members of the family to the older; and it often reconstructed member languages through the postulation of regular diachronic "laws" of linguistic change. The acceptance of such a goal presupposed a model of the nature of linguistic change and of the evolution of new language varieties from older ones. One such a model--only implied in the early writings of the tradition, but fully spelled out by the end of the nineteenth century--is referred to as the Stammbaum model.[13] Although originally intended to serve as a frame of reference for linguistic reconstruction, this model has been highly influential in work carried on in language classification in general. It has appeared in one form or another in the classification of many South Asian languages (cf. 2.6.).

The Stammbaum, or "family tree," model of linguistic history was designed both in response to and as a legitimization of the attempted reconstruction of proto-Indo-European through the systematic comparison of the morphological and phonological forms of its various daughter (i.e. contemporary) languages.[14] Based on the works of Grimm, Rask, Sir William Jones and others, and first expounded by Bopp in the mid-nineteenth century, *the comparative method* of linguistic reconstruction was used to determine the forms of earlier stages of languages through a systematic comparison of forms in the presumably related contemporary varieties of languages.[15] The basic method is to postulate ancestral forms (marked with the symbol *) which, after the operation of well-motivated historical rules,[16] result in the descendant forms in each of the offshoot languages. Since the reconstruction draws its legitimacy from the overall simplicity of the reconstructed system, in some cases the postulated forms may be identical to one of the alternating descendant forms, while in other cases it may not closely resemble any of them. Ideally, the earlier (or "proto") form of a language should be ascertainable from a systematic internal analysis of its various derivative forms. If written records of one or more of the earlier stages of a language exist, they can serve as a check on the veracity of the reconstruction.[17] Reconstructed proto-forms are themselves subject to internal reconstruction, with their comparison yielding further proto-forms which are yet another stage removed from the primary base of current language use. In the more simplistic versions of these theories, the diversity of numerous descendant languages is reconstructed back through stages of increasing unity, until a level of complete homogeneity is reached.[18] The best known application of such a model, that of the reconstruction of Proto-Indo-European, was thought by many to lead to the original language of the Indo-European tribes, and a considerable literature was devoted to determining the location ("Urheimat") of the speakers of this language, ascertaining their racial and cultural characteristics, and even composing folk tales in the reconstructed proto-language.[19]

The application of the comparative method of linguistic reconstruction leads to the production of a type of inverted tree structure; a substratum of spoken contemporary languages are said to be derived from a smaller set of feeding branch languages, which are in turn derived from a still smaller set, and ultimately from a single source language. In a simple classification of languages, the members of a family can be progressively subdivided into groups, where membership within a group or subgroup means that the languages are derived from a common reconstructed source. A characteristic example of such a taxonomy is given in Figure 7.

The structure in Figure 7 well typifies the Stammbaum model. Essentially, primary data exists for only the currently spoken varieties (bottom level). Approximations of the structures of all other varieties must be reconstructed by comparison of the structures of the bottom-line

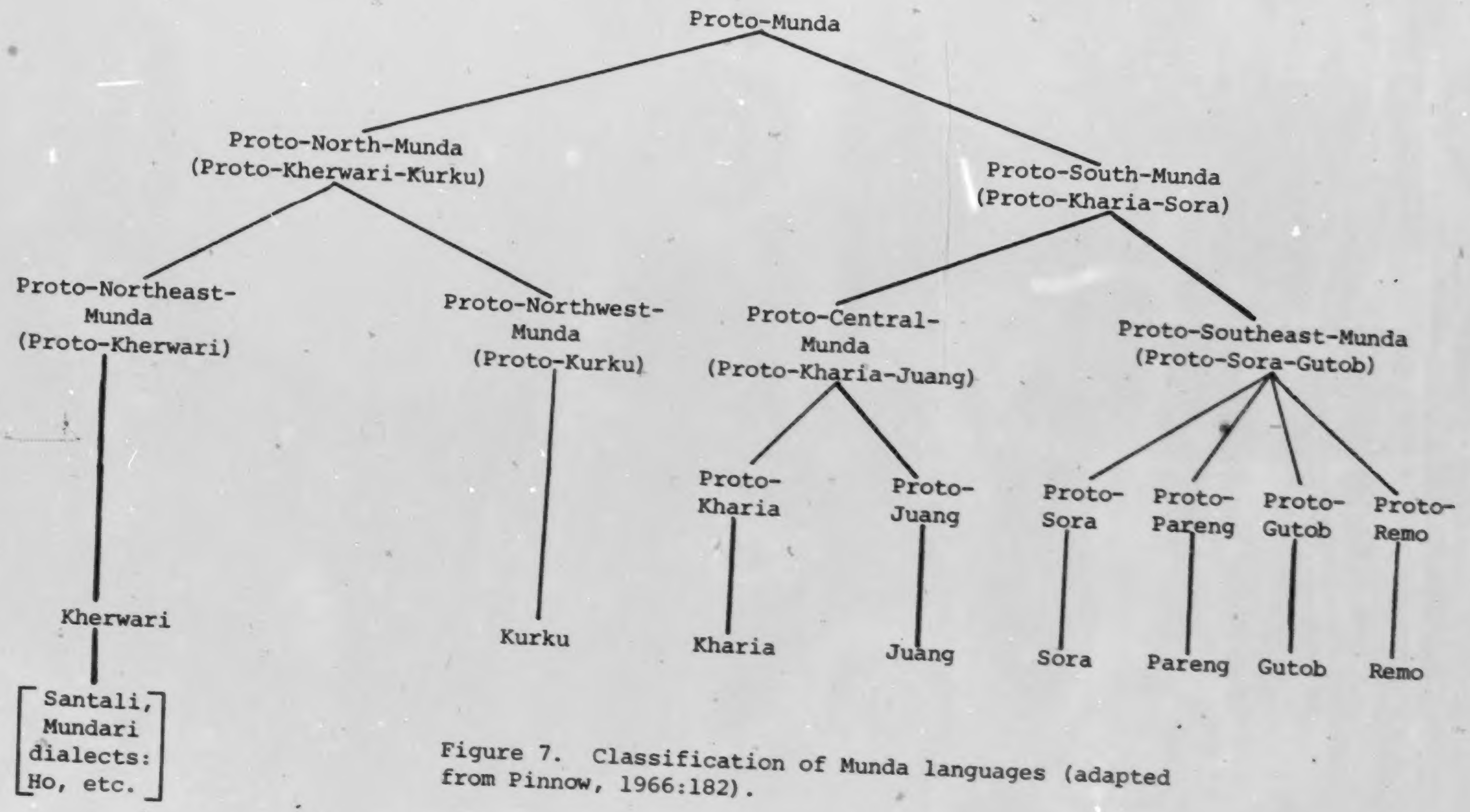

Proto-Munda
Proto-North-Munda (Proto-Kherwari-Kurku)
Proto-South-Munda (Proto-Kharia-Sora)
Proto-Northeast-Munda (Proto-Kherwari)
Proto-Northwest-Munda (Proto-Kurku)
Proto-Central-Munda (Proto-Kharia-Juang)
Proto-Southeast-Munda (Proto-Sora-Gutob)
Proto-Kharia
Proto-Juang
Proto-Sora
Proto-Pareng
Proto-Gutob
Proto-Remo
Kherwari
Kurku
Kharia
Juang
Sora
Pareng
Gutob
Remo
Santali, Mundari dialects: Ho, etc.

Figure 7. Classification of Munda languages (adapted from Pinnow, 1966:182).

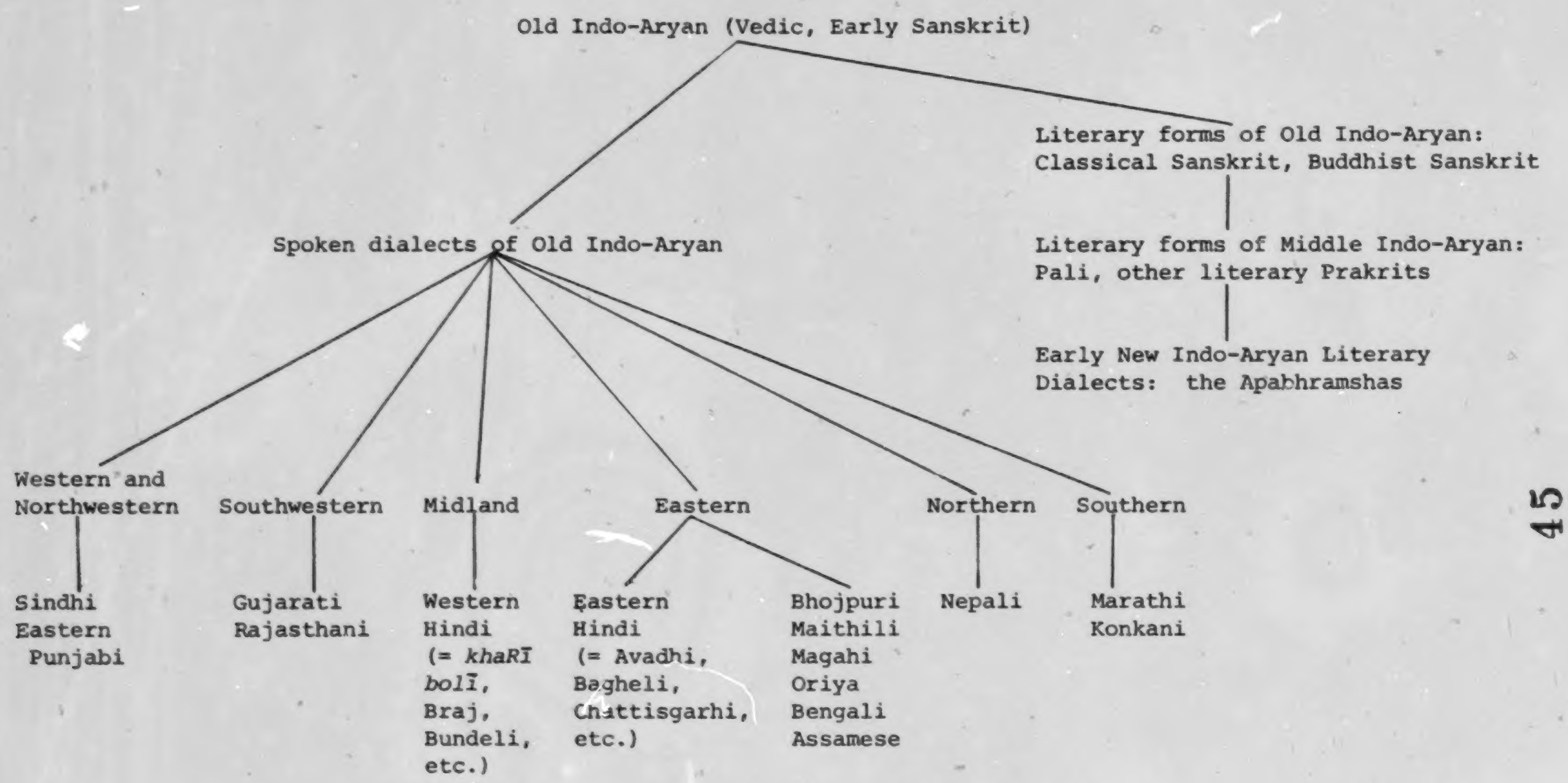

Old Indo-Aryan (Vedic, Early Sanskrit)
Literary forms of Old Indo-Aryan: Classical Sanskrit, Buddhist Sanskrit
Literary forms of Middle Indo-Aryan: Pali, other literary Prakrits
Early New Indo-Aryan Literary Dialects: the Apabhramshas
Spoken dialects of Old Indo-Aryan
Western and Northwestern
Southwestern
Midland
Eastern
Northern
Southern
Sindhi
Eastern Punjabi
Gujarati
Rajasthani
Western Hindi (= *kharī bolī*, Braj, Bundeli, etc.)
Eastern Hindi (= Avadhi, Bagheli, Chattisgarhi, etc.)
Bhojpuri
Maithili
Magahi
Oriya
Bengali
Assamese
Nepali
Marathi
Konkani

Figure 8. Stammbaum of Indo-Aryan languages.

language varieties. In Figure 7, each level in the diagram represents a generation in the family history of the Munda languages. The level of Proto-Northeast-Munda (Proto-Kherwari) is reconstructed from an internal analysis of the structures of the Kherwari and Santali subgroups of contemporary spoken forms. Proto-North-Munda is a reconstructed language arising from a comparison of the reconstructed Proto-Northeast-Munda and Proto-Northwest-Munda.

The family trees resulting from comparative analysis can be of essentially two sorts. The first, which we shall call *historical,* represents the reconstructed stages in the history of a language family. The entries at the bottom level of such a diagram most closely resemble modern spoken varieties,[20] and those at the top indicate more distantly removed stages, which are arrived at only by reconstruction (there often being no written records of these earlier periods). Such charts frequently have entries for both reconstructed stages and stages for which there are extant records. The family tree history of the Indo-Aryan languages given in Figure 8 illustrates this type of situation.

In contrast are what can be called *typological* charts, in which the bottom levels of the tree represent contemporary varieties of languages and the upper levels progressively larger groupings of languages. Each level represents an order of related languages, with all languages grouped together presumably derived from a common ancestor. Another chart of the Munda languages by Pinnow (Figure 9) is of this sort.

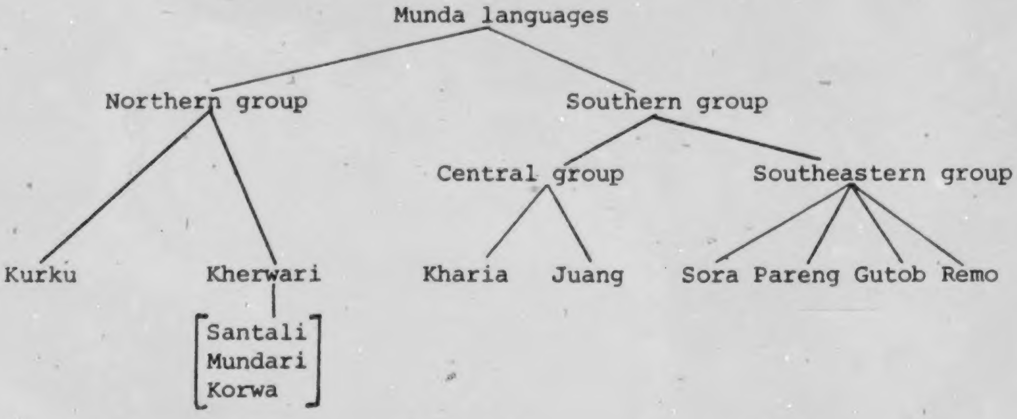


Figure 9. The Munda languages arranged typologically
(from Pinnow, 1963)

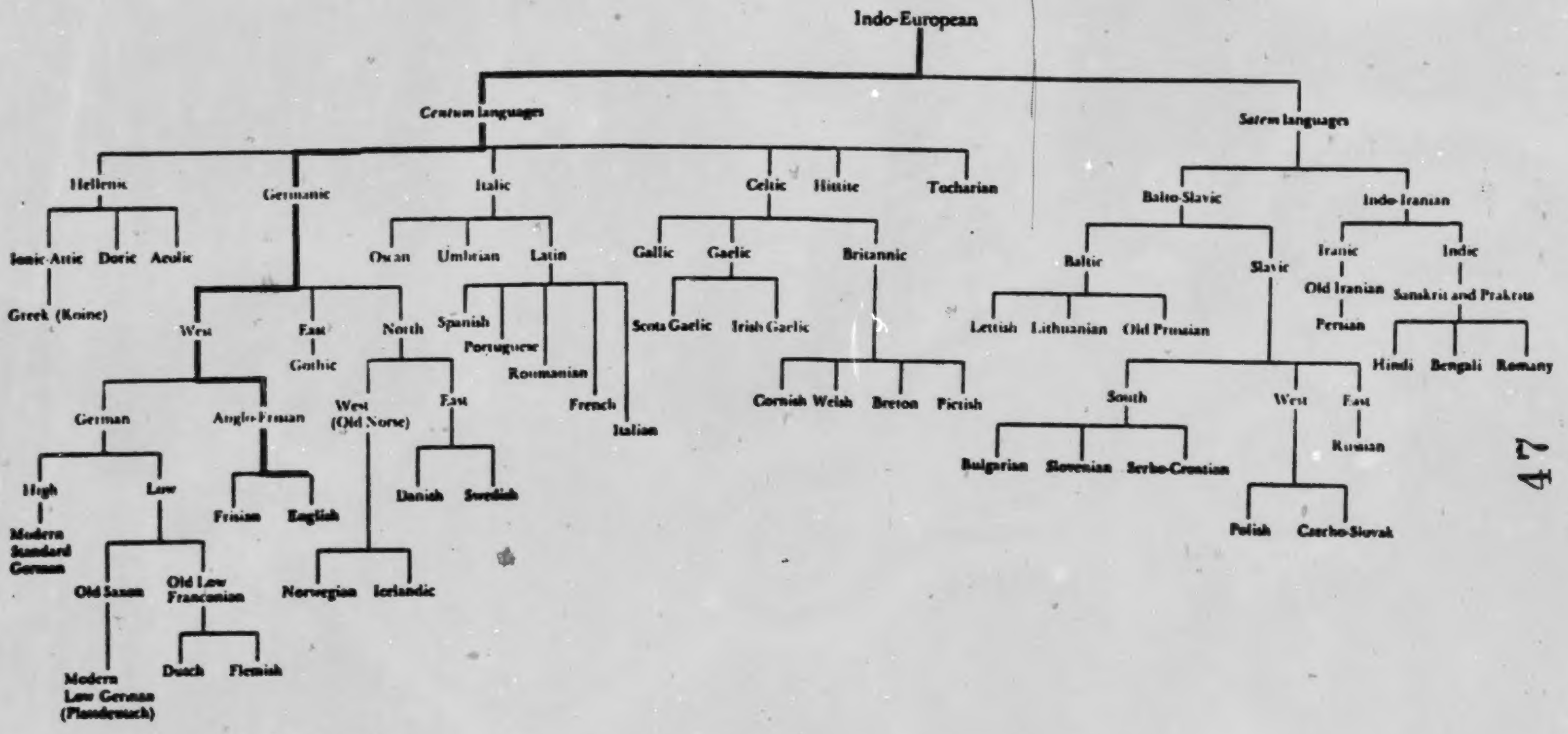

Indo-European
Centum languages
Satem languages
Hellenic
Ionic-Attic
Doric
Aeolic
Greek (Koine)
Germanic
West
East
Gothic
North
German
Anglo-Frisian
High
Modern Standard German
Low
Old Saxon
Modern Low German (Plattdeutsch)
Old Low Franconian
Dutch
Flemish
Frisian
English
West (Old Norse)
Norwegian
Icelandic
East
Danish
Swedish
Italic
Oscan
Umbrian
Latin
Spanish
Portuguese
Roumanian
French
Italian
Celtic
Gallic
Gaelic
Scots Gaelic
Irish Gaelic
Britannic
Cornish
Welsh
Breton
Pictish
Hittite
Tocharian
Balto-Slavic
Baltic
Lettish
Lithuanian
Old Prussian
Slavic
South
Bulgarian
Slovenian
Serbo-Croatian
West
Polish
Czecho-Slovak
East
Russian
Indo-Iranian
Iranic
Old Iranian
Persian
Indic
Sanskrit and Prakrits
Hindi
Bengali
Romany

Figure 10. Stammbaum of Indo-European languages (from Pyles, 1964:80-1).

Technically speaking, the two types of maps just discussed need to be kept distinct, although in practice they are often not. It is not uncommon to find typological and historical classifications mixed within the same Stammbaum. The diagram of the Indo-European languages shown in Figure 10 is of this sort. Many of the node labels such as *Centum* languages, *Satem* languages, Germanic, High German, etc. represent broad groupings of languages, while others such as Old Iranian, Old Prussian, etc. indicate historical stages in the development of languages.

The essential prerequisite for establishing either typological or historical Stammbaum diagrams is the existence of linguistic correspondences between related speech forms. The comparative linguist looks at presumed cognate linguistic forms in the related languages and postulates earlier forms from which the later ones may most optimally be derived. A simple example of the results of this procedure is given in Figure 11.

	Tagalog	Javanese	Batak	Primitive Indonesian
'choose'	l 'pi:liʔ	l pilik	l pili	l *pilik
'lack'	l 'ku:laŋ	r kuraŋ	r huraŋ	L *kuLaŋ
'nose'	l i'lu	r iru	g igu	_g_ *i_g_u
'desire'	l 'hi:lam	D iDam	d idam	D *hiDam
'point out'	r 'tu:ruʔ	d tuduk	d tudu	d *tuduk
'spur'	r 'ta:ri	_d_ ta_d_i	_d_ ta_d_i	_d_ *ta_d_i
'sago'	g 'sa:gu	g sagu	g sagu	g *_t_agu
'addled'	g bu'guk	Ø vuʔ	r buruk	γ *buγuk

Figure 11. Reconstruction of Primitive Indonesian consonants (adapted from Bloomfield, 1933:310).

Each of the reconstructed Primitive Indonesian forms is obtained by comparison of the members of a set of alternating lexical items. As such there is little external validation of the resulting reconstruction; however, such reconstructions attain legitimacy when it is possible to demonstrate systematic phonological or morphological correspondences holding between the members of many sets of words.[21]

A very simple example of such a correspondence is demonstrated by the observation that initial *p* in numerous Latin lexical items corresponds to *f* in Germanic forms having the same gloss, and to zero in the Celtic languages (e.g. Latin *pater;* English *father;* Old Irish *'aδir.*) The efficacy of postulating a hypothetical base form beginning with *p* is demonstrated by the Sanskrit forms *pitā* and Greek *pātēr*. The correspondence can be extended to such sets as Latin *porkus* ('pig'), Old English *fearh* (modern English *farrow*) and Old Irish *ork;* Latin *penta,* English *five.* Properly speaking these comparisons should be made only between historically comparable stages of the languages. Thus it is necessary to compare Latin with Classical (and not Modern) Greek, Old English, and some form of Old Celtic.

Once this kind of reconstruction has been performed, it is possible to postulate sound laws which relate the different stages in the history of a language family. In the example just given, a sound law (known as the First Germanic Consonant Shift or "Grimm's Law") relates the Germanic *f* in *father* with the reconstructed **p* from which it is said to be derived. (Note that the rules do not derive *f* from the *p* in *pater,* but rather from a reconstructed "proto-*p*" which is historically prior to either the *p* in Latin *pater* or the *f* in English *father.*) In this case only a single rule *p* → *f* is needed to explain the correspondence, while in others a sequence of rules may be necessary. Thus, for example, it is necessary to postulate two stages in the development of Germanic *d* from Proto-Indo-European **dh* (corresponding to Latin *f* and Greek *th*). In the first of these changes, the voiced aspirate *dh* changes into the voiced spirant δ, which in turn changes into the voiced lenis stop *d*. (e.g. Sanskrit *'a-dhāt* 'he put'; Greek *'thēsō* 'I shall put'; Latin *fēcī* 'I made, did'; English *do.*)

Of greater significance than simple lexical correspondences for internal reconstruction are instances in which historical rules postulated for relating different stages of a language family have structural affinities enabling them to be collapsed under single general rules. Thus for example, three rules which state that Indo-European *b* → *p*, *d* → *t*, and *g* → *k*, might be collapsed with a general rule that states that voiced stops become devoiced. The greater the generality of the rules relating language varieties, the more systematic the structural relationship between the stages or competing forms.

In the most simplistic versions of the Stammbaum theory of language history it is claimed that related languages can be reconstructed back to a homogeneous common source. However, this claim came under serious attack quite early in the tradition of Indo-European comparative grammar. First of all it was argued that there is no reason to assume that the various daughter offshoots of a common historical source all split off at the same time. It is more reasonable to suggest that geographical and social groups are constantly integrated into or separated from main linguistic groups, and that linguistic change, rather than occurring in discrete stages, is an ongoing process. If this is true then language charts such as that shown in Figure 10 might better be replaced by those of the sort in Figure 12.

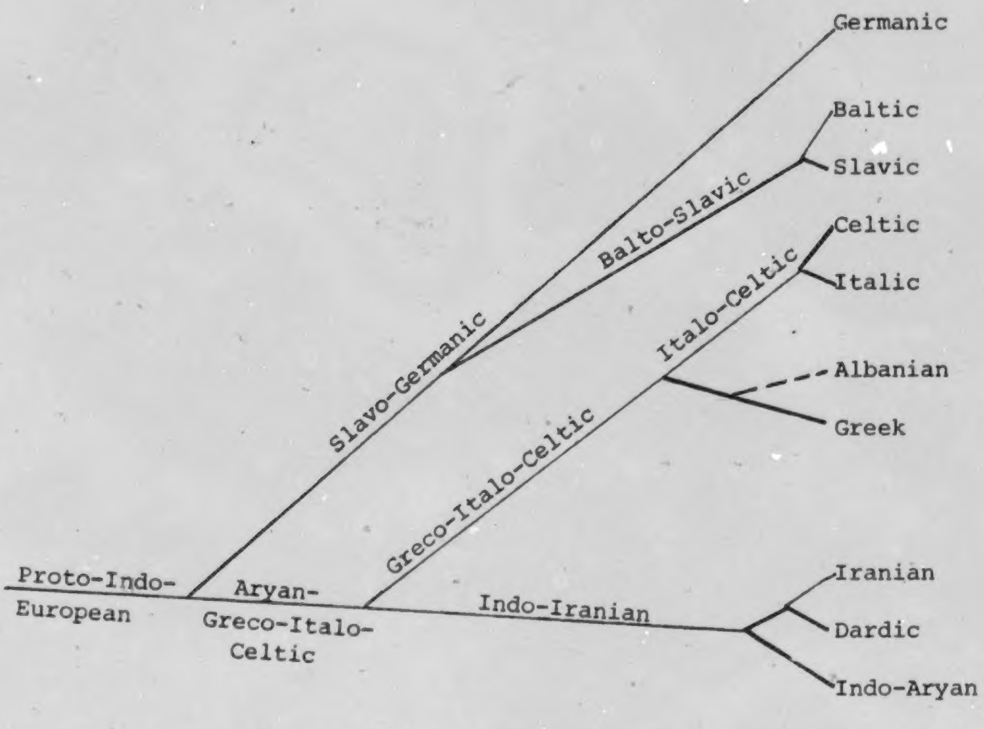

Germanic
Baltic
Slavic
Balto-Slavic
Celtic
Italic
Italo-Celtic
Slavo-Germanic
Albanian
Greek
Greco-Italo-Celtic
Proto-Indo-European
Aryan-Greco-Italo-Celtic
Indo-Iranian
Iranian
Dardic
Indo-Aryan

Figure 12. Stammbaum of the Indo-European languages, version 2. (Adapted from Schleicher as cited in Lehmann, 1962:139).[22]

In spite of their not inconsiderable differences, the diagrams in Figures 10 and 12 are alike in postulating a single Proto-Indo-European from which all of the Indo-European languages are derived. Other models postulated within the field of comparative Indo-European questioned the reconstructibility of these languages back to a single source.[23] It has been claimed, with good reason, that a systematic examination of the Indo-European languages will yield a number of areas in which it is impossible to reconstruct a single precursor language.[24] If this is so, then Proto-Indo-European must have consisted not of a single monolithic language, but of a number of related geographically distributed "dialects." This view maintains that linguistic diversity is not necessarily increased through time, and that dialect forms are as likely to exist in the early stages of a language family as in its later ones. The classification of Indo-European languages, according to such a model, might then look like the structure in Figure 13.

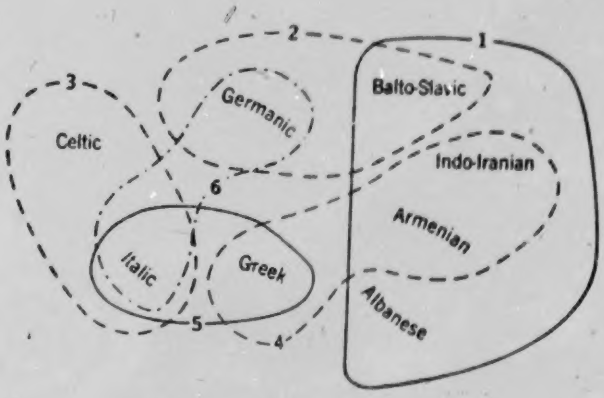


Figure 13. "Overlapping" diagram of Indo-European family of languages. (Adapted from Schrader and cited in Bloomfield, 1933:316).

1. Sibilants for velars in certain forms.
2. Case-endings with [m] for [bh].
3. Passive-voice endings with [r].
4. Prefix ['e-] in past tenses.
5. Feminine nouns with masculine suffixes.
6. Perfect tense used as general past tense.

The disparity between these three approaches to language classification, which we can perhaps call pure-Stammbaum, modified Stammbaum, and non-Stammbaum, is related to conflicting views on the nature of linguistic change. The pure Stammbaum approach presupposes the existence of a homogeneous speech code. If a subgroup of the original community breaks off from the main group, the density of interaction between the members of the groups will decrease;[25] consequently the likelihood increases that the speech varieties of the independent groups will display innovations. These innovations may start as microscopic alternations in the phonetic makeup of one or more members of the sound system of the innovative language variety, and gradually produce a distinct phonological unit. Once these changes have occurred, they spread throughout the structural system through analogic change.

Frequently such extensions of the domain of linguistic innovation bring units of a linguistic system into conflict with other units or structural properties. The development of such areas of conflict is often resolved by further linguistic change. The instigation of even a small phonetic alternation can trigger a chain of linguistic innovations, the result being considerable linguistic change.[26] Such innovations are as likely to occur in the group of speakers branching off as in the original body of speakers. The results of changes occurring in each of these groups compound and in time lead to the emergence of distinct dialects, which, through political and social recognition, eventually are accorded the status of distinct languages.[27] The modified Stammbaum approach is essentially identical to this, except that it allows the breaking off of groups to be staggered through time, and correlates this branching with the migrational patterns of distinct ethnic groups.[28]

Clearly in opposition to the views on sound change incorporated into the Stammbaum theory is a view (most commonly attributed to J. Schmidt, a student of August Schleicher) which admits the possibility that linguistic innovation can occur frequently in speech forms that are not fully cut off from larger linguistic communities; that is, that innovation can occur within groups which are still in cultural and geographical contact with one another.[29] This theory denies that sound change necessarily produces the sharp cleavage of the sort demonstrated in the Stammbaum diagrams. Rather, Schmidt and others claimed that "the splitting process begins subdialectally and proceeds through increasing dialectal divergence until the assumption of two or more distinct languages is warranted" (Robins, 1967:179).[30] This process is said to be a lengthy one, and until such time that complete cleavage occurs, a good deal of cultural and linguistic contact between the speakers of the distinct varieties must be assumed. It is not claimed that all instances of linguistic innovation are of this sort, but only that in many cases of language change this continued contact is present.[31]

According to this model, language change occurs not because of a sudden innovation prompting a series of linguistic alternations leading to the rapid emergence of distinct dialects; rather because, as in Leonard Bloomfield's paraphrase of the model,

> linguistic changes may spread, like waves, over a speech-area, and each change may be carried out over a part of the area that does not coincide with the part covered by an earlier change. The result of successive waves will be a network of isoglosses. Adjacent districts will resemble each other most; in whatever direction one travels, differences will increase with distance, as one crosses more and more isogloss-lines. . . . Now, let us suppose that among a series of adjacent dialects, which, to consider only one dimension, we shall designate as A, B, C, D, E, F, G, . . .X, one dialect, say F, gains a political, commercial, or other predominance of some sort, so that its neighbors in either direction, first E and G, then D and H, and then even C and I, J, K, give up their peculiarities and in time come to speak only the central dialect F. When this has happened, F borders on B and L, dialects from which it differs sharply enough to produce clear-cut language boundaries; yet the resemblance between F and B will be greater than that between F and A, and, similarly, among L, M, N, . . .X, the dialects nearest to F will show a greater resemblance to F, in spite of the clearly marked boundary. . . . (Bloomfield, 1933:317-18).

The application of this theory of sound change, called the wave theory (*Wellentheorie*), can be seen in the distribution of isoglosses in Figure 2 above. A classic application is in the description of the so-called "Rhenish fan," a set of four isoglosses whose geographical distribution in Germany resembles a fan (Figure 15). Each of these isoglosses represents an alternation with regard to a significant variable, the alternations resulting from a series of ordered phonological changes which occurred in the evolution of Modern German. These changes are summarized in Figure 14. The changes did not occur, however, without exception throughout the Old High German dialect area; rather they were sealed within the area. This pattern of distribution accounts for the spreading of the "fan" as shown in Figure 15.

1. Late Proto-Germanic [p-, t-, k-, -pp-, -tt-, -kk-] → Old High German [pf, ts, k(x)]

   E. *pool:* Ge. *Pfuhl*
   E. *tongue:* Ge. *Zunge*
   E. *cow:* Ge. *Kuh,* but Swiss *kxū*
   E. *shape:* Ge. *schöpfen*
   E. *sit:* Ge. *sitzen*
   E. *wake:* Ge. *wecken,* Swiss Ge. *wekxen*

2. Late Proto-Germanic [-p-, -t-, -k-, -p, -t, -k] → Old High German [-f(f), -s(s), -x(x)]

   E. *hope:* Ge. *hoffen*
   E. *water:* Ge. *Wasser*
   E. *cake:* Ge. *Kuchen*
   E. *up:* Ge. *auf*
   E. *it:* Ge. *es*
   E. *book:* Ge. *Buch*

Figure 14: Germanic consonant shifts (data from Lehmann 1962:122).

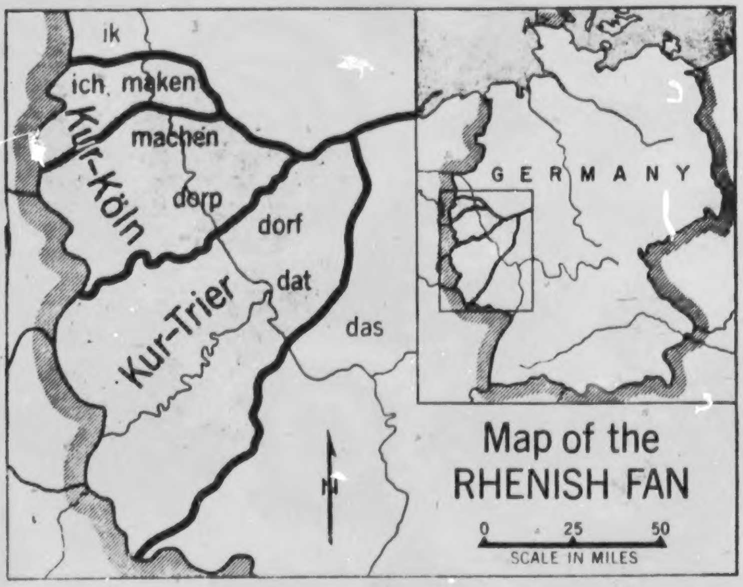

ik
ich
maken
machen
Kur-Köln
dorp
dorf
dat
Kur-Trier
das
N
GERMANY
Map of the RHENISH FAN
0
25
50
SCALE IN MILES

Figure 15: Map of the "Rhenish fan" (from Lehmann, 1962:124).

This pattern of diverging isoglosses clearly was problematic for adherents of the conventional Stammbaum theory. The wave theory was developed in part as a response to the inability of the Stammbaum theory to explain speech continua--that is, areas in which dialects blend imperceptibly into one another, and where there are no clearly marked language or dialect boundaries--and was intended to supplement rather than replace the Stammbaum model. The wave theory, by maintaining that sound changes can be disseminated from central foci, thus creating an effect which resembles the overlapping of waves, facilitated the description of speech areas lacking well defined bundling of isoglosses.

In recent years there has been a significant interest in the formal mechanisms by which sound changes are disseminated once initiated, i.e. in the mechanism for the spread of the "wave".[32] Perhaps the most significant contribution in this area was made by C.-J. Bailey who has developed what is referred to as the "new wave model" (Bailey, 1972; 1973a). Bailey's major advance lies in his attention to the output of rules, both original phonological rules and the innovations which replace them. A rule which has not yet applied obviously has no output; one which has been fully assimilated has a categorical output; and a rule which is in the process of change has a variable output--that is, there are some contexts in which it applies and others in which it does not. The wave-like spread of a set of sound changes can be graphically represented by charting the type of output each of a number of sound rules has at a finite number of points in time. Using the symbols 0, x, and 1 to refer respectively to zero output, variable output, and categorical output of phonological rules, the spread of four phonological rules can be diagrammed as in Figure 16. The pattern displayed at Time vii in many ways corresponds to the situations shown on the map of the "Rhenish fan"; it characterizes a set of overlapping dialects, with isoglosses indicating the spreading out in space of particular phonological features. Such a situation is tantamount to a speech continuum. (This schema is valid only when the spread of a linguistic innovation is unrestricted in all directions. A wave spread checked in one direction appears as in Figure 17.)

The Stammbaum theory and the wave theory (in both of its versions) are models designed to account for the notion of "historical relatedness" among language varieties. As such they make claims both about the opposition between language and dialect and about how the linguistic differences which characterize this opposition occur and are transmitted in time and space. That they make conflicting claims about the nature

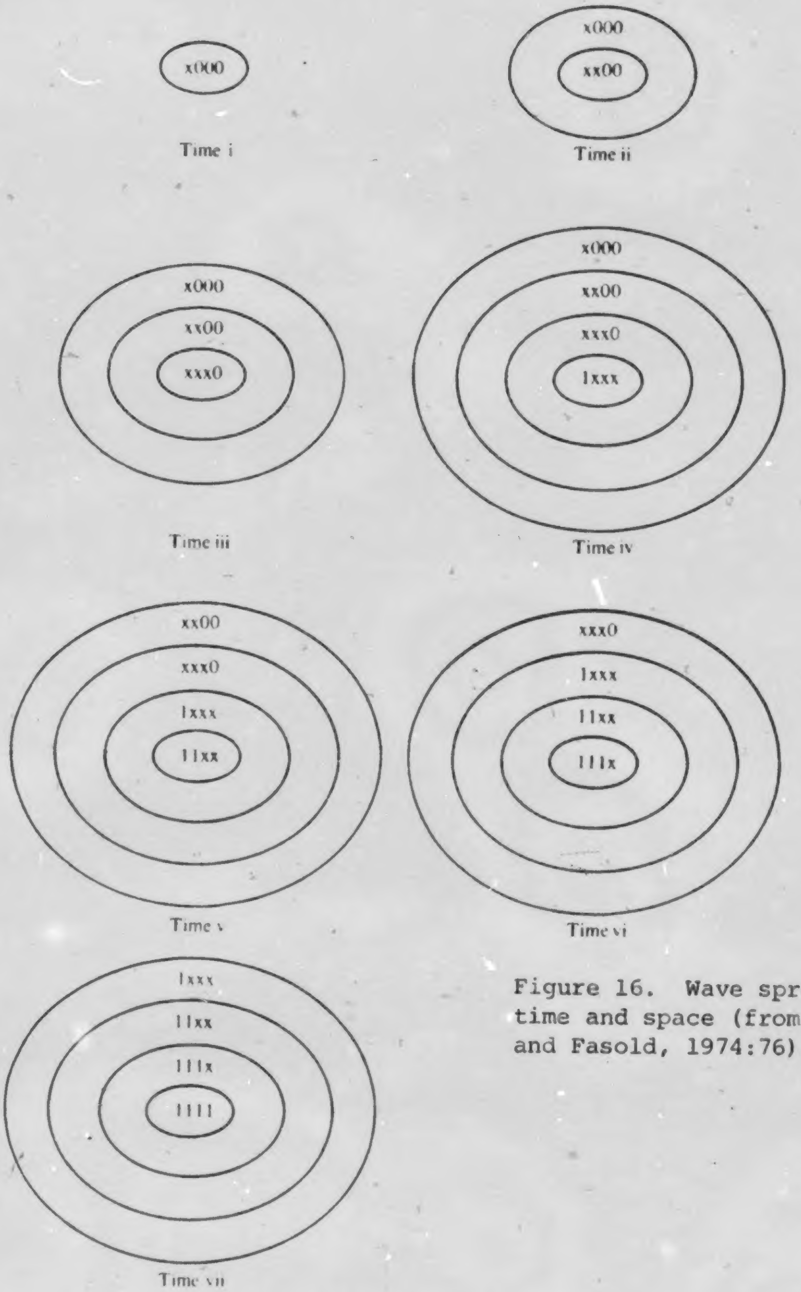

x000
Time i
x000
xx00
Time ii
x000
xx00
xxx0
Time iii
x000
xx00
xxx0
1xxx
Time iv
xx00
xxx0
1xxx
11xx
Time v
xxx0
1xxx
11xx
111x
Time vi
1xxx
11xx
111x
1111
Time vii

Figure 16. Wave spread in time and space (from Wolfrom and Fasold, 1974:76).

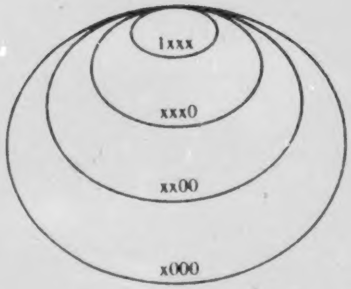


Figure 17. Wave with spread arrested in one direction (from Wolfram and Fasold, 1974:77).

of language change may not reflect on their adequacy or inadequacy as theories per se, but rather on their suitability for describing actual situations of linguistic diversity. The wave model was designed to compensate for the inability of the Stammbaum to handle speech continua, as are found along the Franco-Italian border, in the "Rhenish fan" area, and across most of North India.[33] As such, and particularly as modified by scholars such as Bailey, it provides one means of describing a "substratum" level of language use, one differing from such notions as "standard language" and "normative variety of language." The Stammbaum model, by contrast, is best suited for characterizing the development of "standard languages," by a systematic disregard of the variable output of innovative rules, and by making symmetrical the time factors involved in linguistic innovation.

2.3. Structural Dialectological Models of Linguistic Relatedness

In earlier sections of this chapter we examined the claim that isoglossal theories of language and dialect as well as the Stammbaum approach to linguistic reconstruction are useful models for characterizing the historical and synchronic connection of language varieties whose essential relatedness can be presupposed. We also asserted that one of the major weaknesses of these theories is that they provide little means for formally characterizing the relatedness of systems when considered as a whole, and by extension, of making decisions about the relatedness of speech varieties whose mutual affinity is open to question.

The development of "structuralist" movements within linguistics at the beginning of the twentieth century[34] led to a reexamination of the methodological bases of language classification, which in turn exposed some of the weaknesses of earlier models. At the beginning of the twentieth century, Saussure and others had convincingly shown that linguistic systems can be conceptualized as abstract structured entities, and that each unit of the system is not independent, but rather draws its functional significance from its relationships with all other elements in the system.[35] To Saussure, linguistic systems are analogous to what Durkheim called "social facts"--aspects of society which are characterizable as tightly woven systems. It makes little sense to talk of phonemes, morphemes, and the like, without making reference to an overall conception of language within which these units exist and function. The totality of a system enables the isolation of its components.[36]

The Saussurian insistence on treating language as a fully integrated structure was directed primarily against some aspects of Neogrammarian theory and against many practices of the early dialect geographers.[37] The structuralists' objections to the Neogrammarians focused on the latter's attitude toward sound laws. For many Neogrammarians the postulation of categorical sound laws to relate synchronically observed correspondences between cognate linguistic forms was the major objective of linguistic inquiry. The asserted historical validity of these sound laws formed the major basis for the Neogrammarians' claims of the scientific nature of their work.[38] Yet to the early structuralist it was not sufficient to state that sound p in language P corresponds to sound q in language Q, and that this fact can be explained by deriving both p and q from the reconstructed form $*r$, and by postulating rules $r \rightarrow p$ in P and $r \rightarrow q$ in Q. They asserted rather that in order to understand the sound change fully it is necessary to determine the functional status of $*r$ in the reconstructed proto-system as well as that of p within P and q within Q.[39] Historical relatedness could no longer be asserted on the basis of isolated correspondences, but had to be verified by systematic comparison of whole languages.

The early structuralist criticism of many late 19th and early 20th century dialect geographers was of a similar sort. Structuralists repeatedly charged that the map construction of many of the dialect geographers was but an antiquarian game of little linguistic significance.[40] True generalizations about dialects can only be derived through examination of total linguistic systems. Plotting the various lexical realizations of single items tells little of how these items function in the systems from which they are extracted. The extreme position maintained by Bartoli and the so-called "Neolinguists"--that borrowing, innovation, and cross-cultural linguistic contact are so pervasive in the real world as to prohibit the formulation of any categorical phonological rules, and to require that each word (and presumably each structural unit of a language) be conceived as an isolated entity having its own history,[41] and describable purely on its own terms--was even more untenable.

The term "structural dialectology" has come to refer to the collective attempts of a number of linguists to apply models of structural linguistics to the problems of language distribution and classification.[42] Such attempts presuppose that the comparison of language varieties proceeds from a discussion of linguistic systems as wholes, and is not restricted to select subsystems. The arguments advanced concerning how such a systematic comparison takes place, and what is meant by a linguistic system, vary among structuralist schools. In its general thrust, however, structural dialectology is essentially an attempt to make the results of dialectology compatible with those of general linguistics. The adequacy of particular models is no greater than the adequacy of the models of language description presupposed by the adherents of structural dialectology.

A number of questions need to be raised about the above assertions concerning dialectological investigation. How is one to describe the relationship holding between formal linguistic "elements" of a single dialect, and how can the relationship between dialects be characterized in terms of these elements? In answer to these questions most early structuralist models of language either stated or assumed that a linguistic system contains fundamental units at each of a number of levels, with each unit standing in theoretical opposition to all other units at the same level of analysis. Relations between units at the same level can be thought of as either paradigmatic (sometimes called relations *in absentia*, and which refer to the abstract interconnections between units, totally apart from the use of these units in particular collocations) or syntagmatic (also referred to as relations *in praesentia*, and concerned with the connections between units as they are distributed in the set of morphological and lexical collocations of the language).[43] In addition to making a distinction between syntagmatic and paradigmatic relations, most structuralists maintain that it is necessary to make a distinction between synchronic and diachronic analyses of language varieties, and that adequate synchronic description of a language can and should be carried on without regard to diachronic considerations.[44] The comparison of related language varieties must proceed by an analysis of synchronically adequate descriptions of the individual varieties. Single elements of grammatical systems can only be compared in the context of the matrix of paradigmatic and syntagmatic oppositions in which those elements synchronically enter.

These positions are perhaps nowhere better observed than in a paper by Trubetzkoy entitled "Phonologie und Sprachgeographie" in which the author attempts to provide a theoretical framework for comparing grammatical systems. For Trubetzkoy, as well as for most other early structural linguists, the most accessible aspect of language for systematic comparison is phonology. He attempts to set out the basis for cross-dialectal comparison in terms of abstract configurations of underlying elements as well as in terms of the distribution and functional uses of these elements. Trubetzkoy states that:

> Die lautlichen Unterschiede zwischen zwei Dialekten können dreifacher Art sein: sie können das *phonologische System* betreffen oder die *phonetische Realisierung* einzelner Phoneme oder die *etymologische Verteilung* der Phoneme in den Wörtern. Demnach reden wir von *phonologischen*, *phonetischen* und *etymologischen* Dialektunterschieden. (Trubetskoy, 1958:262).

Dialects can be distinguished not only by the number and intrinsic content of their phonemes, but also by the distribution of these elements:

> Die *phonologischen* Dialektunterschiede zerfallen wiederum in *Inventar-* und *Funktionsunterschiede*. Ein phonologischer Inventarunterschied besteht, wenn der eine Dialekt ein Phonem besitzt, das einem anderen Dialekte unbekannt ist. Ein phonologischer Funktionsunterschied besteht, wenn ein Phonem in dem einen Dialekte in einer phonologischen Stellung vorkommt, in der es in einem anderen Dialekte nicht vorkommt. (Trubetzkoy, 1958:262).

By differentiating between functional and intrinsic content of phonological systems, Trubetzkoy is able to superimpose discreteness onto the continuum of purely phonetic representations. The comparison of entire linguistic systems must be done in terms of systems within which elements can be isolated in terms of their functional opposition to each other.

The importance of the function of elements in a phonological description has been commented upon by Uriel Weinreich in his important paper, "Is a Structural Dialectology Possible?" In discussing the functional importance of elements Weinreich cites the hypothetical case of a language in which four speakers realize the word *man* as [man], [man], [mån], and [mån] respectively. In this imaginary language, Speaker 1 functionally distinguishes vowel length; his rendition of *man* is then phonemically $_1$/măn/. Speaker 2's idiolect does not distinguish among vowel lengths and his version of *man* can be phonemicized as $_2$/man/. Speaker 3's *a*-like vowel has the allophonic variant [å] between /m/ and /n/. His [å] is thus a contextually determined allophone of /a/. Speaker 4's idiolect doesn't display the allophonic variation seen in Speaker 3; his [å] is most likely a manifestation of an independent phoneme /o/.

There are a number of ways to represent the distinctions displayed in the four realizations of the vowel in *man*. To the pre-structuralist dialectologist the opposition can be represented by drawing a single isogloss across an imaginary space, the isogloss marking the limits of the occurrence of the phonetic alternates [a] and [å] (Figure 18).

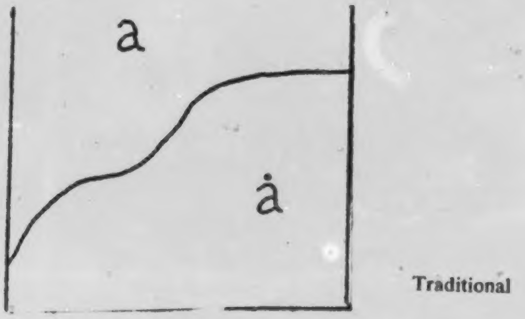


Figure 18. The vowel in *man* in language X (from Weinreich, 1954:311).

For Weinreich, however, an accurate characterization of the dialect makeup of language X requires charting the phonemic membership of each occurring phonetic element. This can be represented as shown in Figure 19.

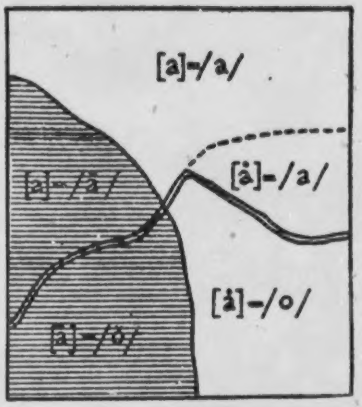


Figure 19. Structural analysis of dialect structure of language X based on vowel in *man*.

A continuous single line divides areas with different phonemic inventories (shaded area distinguishing vowel length, unshaded area not distinguishing it). The double lines separate areas using different phonemes in the sample word (difference of distribution). The dotted line separates allophonic differences (Weinreich, 1954:311).

A more complex structural analysis of a dialect area has been presented by Stankiewicz and concerns dialect boundaries in northeastern Poland. Stankiewicz points out that the main criterion for establishing these boundaries has been the historical development of the hushing spirants and affricates **š*, **ž*, **č*, and **ǯ*. The differential development of these sounds, and the dialect boundaries established with respect to it, are shown in Figure 20.

The phonological information represented in Figure 20 is summarized by Stankiewicz as follows:

> In the blank areas **š*, **ž*, **č*, **ǯ* coalesced with *s*, *z*, *c*, *ʒ* (the so-called "mazurzenie" dialects); in the areas marked with horizontal lines, **š*, **ž*, **č*, **ǯ* changed into *ś*, *ź*, *ć*, *ʒ́* or fused into a single series *š́*, *ž́*, *č́*, *ǯ́* (the so-called "siakanie" dialects); whereas in the areas marked with slanted lines, the original three series *š*, *ž*, *č*, *ǯ*; *ś*, *ź*, *ć*, *ʒ́*; *s*, *z*, *c*, *ʒ* have remained intact. The "less important"

features have served as the basis for further subdivisions and are indicated on the maps by lines. The area in which *å changed to *a* is delimited on the map by line (c); (g) indicates the area in which voiced consonants occur before vowels and sonorants in final position; (h) marks the southernmost limit of the pronunciation *γara* (Standard Polish /v'ara/); (j) delimits the areas with the pronunciation *śvat* (Standard Polish /śf'at/). (Stankiewicz, 1957:49).

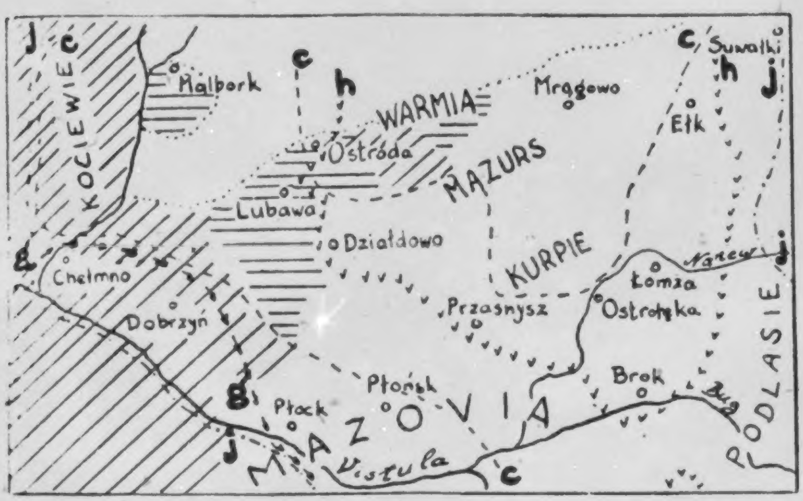

j
c
KOCIEWIE
Malbork
c
h
WARMIA
Mrągowo
c
Suwałki
h
j
Ełk
Ostróda
MAZURS
Lubawa
Działdowo
KURPIE
g
Chełmno
Dobrzyń
Przasnysz
Łomża
Narew
j
Ostrołęka
PODLASIE
Płońsk
Brok
Bug
g
Płock
j
MAZOVIA
Vistula
c

Figure 20. Phonetic boundaries of northeastern Poland (based on S. Urbańczyk, "Zarys dialecktologii polskiej", cited in Stankiewicz, 1957:49).

Stankiewicz asserts that by evaluating the phonological systems of various northeastern Polish dialects in structuralist terms, with reference to the inventory of basic phonological units and their distribution, different dialect demarcations can be constructed. The dialect areas given below are determined on the basis of number and phonological type of phonemes as well as on the basis of the oppositions into which phonemes in particular systems enter. Dialect boundaries determined by these criteria are shown in Figure 21. In the map, dialect Area I encompasses those dialects having seven vowels and twenty-three or twenty-four consonant phonemes. Area II dialect share a higher number of consonants (28) but a reduced number of vowels (5). Areas II and III are distinguished, not by the number of their phonemes, but by the oppositions into which the phonemes enter. The Area III phonemic system shows an opposition between hard and soft velar consonants but lacks the opposition of strident vs. mellow sibilants (š, č, ǯ /ś, ć, ʒ́) found

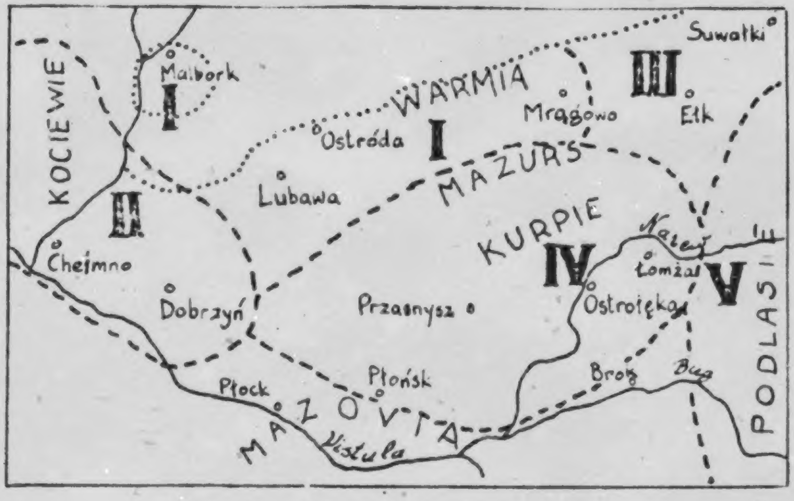

KOCIEWIE
Malbork
I
WARMIA
Ostróda
Mrągowo
III
Suwałki
Ełk
I
MAZURS
Lubawa
II
Chełmno
KURPIE
Narew
Łomża
IV
V
Ostrołęka
Dobrzyń
Przasnysz
Płońsk
Płock
MAZOVIA
Vistula
Brok
Bug
PODLASIE

Figure 21. Phonemic Areas of northeastern Poland (from Stankiewicz, 1957:52).

in Area II. Area IV has twenty-five or twenty-seven consonants and a high number of vowel phonemes (10) which includes both nasals and non-nasals. Area V has seven vowels and thirty consonants, the latter including both hard and soft velars.

The dialect boundaries in northeast Poland shown in Figures 20 and 21 are fundamentally different from each other. In summing up these differences Stankiewicz states the following:

> The historically "important" distinction between the "mazurzenie" and the "siakanie" dialects is of secondary importance to the structuralist, inasmuch as both types have basically two series of spirants and affricates. . . instead of the three series which are found elsewhere. The different "realization" of the hushing consonants in both types is a redundant, not a distinctive phenomenon. The isoglosses (g) and (j) are distributional and lexical. Of "primary" importance to the structuralist are lines (c) and (h), which point to phonemic distinctions. As opposed to the traditional map, with its "primary" and "secondary," i.e. phonetic and distributional features, the structural map is marked by a homogeneity of features and by clearly defined areas. (Stankiewicz, 1957:52).

The features which have been used here in the structural determination of dialect boundaries are phonological ones. The preeminence of phonological criteria for language boundaries has in part been a carry-over from pre-structural historical linguistics, in part a reflection of the views held by some structural schools that higher level analyses of linguistic systems must proceed inductively from analyses of lower-order constituents, and in part a result of the relative accessibility of phonological dialect data. Stankiewicz goes so far as to state that "phonology, which is the most advanced branch of modern linguistics, suggests itself, finally, as the most logical one within which we can determine criteria for the grouping of dialects" (Stankiewicz, 1957:46).

This reliance on phonological criteria is not necessarily characteristic of all schools of structuralist thought, however, and some statements can be found in the literature advocating the possible extension of a structural approach to other aspects of linguistic systems. The incorporation of non-phonological levels of analysis into structural dialectology has had some beneficial spinoffs for language classification in general. Václav Polák (1954) suggests that the difference between "language" and "dialect" might be correlated with structural differences at distinct levels of analysis. He asserts that it is the syntactic and morphological base which distinguishes "languages," while phonological and lexical variations set apart dialects.[45] The efficacy of considering morphological systems has been demonstrated by Joseph Greenberg in a number of papers on the theory of language classification[46] and he has used this approach in drawing up specific taxonomies of the languages of sub-Saharan Africa. The construction of morphologically based dialectal studies requires no major theoretical adjustments, and can easily be achieved through an extension of the concept of synchronic morphology advanced by Roman Jakobson.[47]

In addition to advancing views of the bases for cross-dialectal language comparison, structural dialectologists have concerned themselves with defining orders of language variety which transcend idiolects. The description of such generalized levels of language use is seen as prerequisite to understanding the use of language as a unifying factor within social groups. Weinreich calls such generalized systems "diasystems", and states that dialectology is "the investigation of problems arising when different systems are treated together because of their partial similarity. A specifically structural dialectology would look for the structural consequences of partial differences within a framework of partial similarity" (Weinreich 1954:308). In structural terms, areas of partial similarity or difference between linguistic systems can be specified in terms of the number of units at a given level of description, the distribution of these units, the oppositions which these units enter into at a particular level, and the ways in which these units are either subsumed in or dominate units at higher or lower levels of analysis.

Weinreich gives excellent phonological examples of how partial diasystems can be constructed, and provides notational means for indicating which aspects of a number of related phonological systems are held in common.[48] He suggests that in the hypothetical case of two language varieties with identical five vowel phonemic systems the common phonemic inventory can be represented as

$${}_{1,2}//i \approx a \approx e \approx o \approx u//$$

In the event that one of the dialects substitutes a slightly more open vowel for *e*, the representation of the new diasystem becomes

$${}_{1,2}//i \approx \frac{{}_{1}e}{{}_{2}\varepsilon} \approx a \approx o \approx u//$$

For a case in which one variety has three front vowels where the other has four, the new diasystem is represented as

$${}_{1,2}// \frac{{}_{1}/i \sim e \sim \text{æ}/}{{}_{2}/i \sim e \sim \varepsilon \sim \text{æ}/} \approx a \approx o. . .//$$

Further extending this model, Weinreich asserts that it is possible to present an analysis of a Yiddish vowel "diasystem", when considered as a function of three Yiddish dialects. This analysis is represented as follows:

$${}_{1,2,3}// \begin{array}{c} {}_{1}/i\!: \sim i/ \\ \hline {}_{2}/i \sim I/ \\ \hline {}_{3}\,i \end{array} \approx e \approx \frac{{}_{1}/a\!: \sim a/}{{}_{2,3}\,a} \approx o \approx u //$$

1 = central Polish; 2 = southwestern Ukrainian;
3 = northwestern Ukrainian

Weinreich further asserts that similar differences of inventory of grammatical categories can be allowed within structural dialectology, "e.g. between varieties having two against three genders, three as against four conjugational types, and the like" (Weinreich, 1954:312).

For all of its promise, structural dialectology has failed to produce a rich body of literature dealing with specific language areas of the world. Its main contribution lies in pointing out undesirable consequences of uncritically accepting Neogrammarian and early dialectological views on sound laws and language diffusion. The form of specific structural dialectological descriptions has tended to be a function of the general models of language structure advocated by various schools of structural linguistics. Dialectological studies with the "Pragueian" framework have, in extending Jakobson's and Trubetzkoy's views on phonological and morphological structure, concentrated on the minimal elements which can be correlated with functional oppositions between elements of linguistic systems. In this conception, languages can differ in the number, intrinsic phonological content, and distribution of such "distinctive features."[49] Neo-Bloomfieldian American linguistics, on the other hand, has considered phonemes indivisible units, arrived at through a distributional analysis of a sample of "phones" in a finite corpus of "primary" linguistic data. Oppositions within this framework must be between whole phonemes, and not between any constituents of them.[50]

As linguistic theory has been in a state of flux during much of the mid twentieth century, the application of structuralist theory to dialectological studies was never seriously undertaken on a large scale, and many of the traditional methodological and theoretical assumptions of language which pervaded pre-structuralist dialectology were allowed to go largely unchallenged. Many of the main tenets of structural dialectology, as presented by Weinreich and Stankiewicz, have been incorporated into newer theoretical models of language diffusion which are outgrowths of, and in some cases reactions to, radical changes in linguistic theory which have taken place during the past twenty years. Some of these changes and their applications for dialectology and language classification are discussed below.

2.4. *Transformational Analyses of Dialect Differences*

Within the past fifteen or so years a body of literature has been written within the framework of transformational-generative (T-G) grammar which examines ways of capturing the relationships between partially similar grammatical systems. The solutions advanced within this framework for coping with this problem are necessary outgrowths of philosophical positions and assumptions maintained by transformational grammar, as well as of views concerning the methodology of linguistic investigation. Once again, it must be stated that the rubric of T-G grammar refers to a wide range of theories and methods of dealing with language, and that the views of language to be discussed here may not be congruent with the claims of some individuals. Nevertheless, there are a good many assumptions which are shared by virtually all transformational grammarians, and which show up in most attempts within this school to deal with dialect diversity.[51]

T-G grammar presupposes that "native speakers" of any language possess an internalized system that underlies their ability to produce grammatical and appropriate utterances in a wide range of social contexts. This internalized system is characterized by a set of rules which, when correctly applied, "generate" all and only the set of possible grammatical sentences of the language. (The set of sentences which can be produced by the application of these rules is infinitely greater than the utterances which will actually be produced during the lifetime of the speaker.) The task of linguistics in general is to employ any means to arrive at an approximation of these rules. The rules which are constructed by linguists are in effect predictions about what are grammatical utterances of the language. Wherever the constructed rules enumerate sentences which are at variance with known grammatical sentences of the language, the rules must be changed.

As do other schools of linguistics, T-G grammar postulates a number of abstract levels of analysis, establishes units and principles of organization at each of these levels, and provides principles by which the levels are related to one another to form a coherent system. Different models of T-G grammar make conflicting claims as to the form, structure, and content of these units, rules, principles of organization, etc., but all agree that languages can best be described as multi-leveled structures with formal rules relating the representations of utterances at different levels, and that the totality of such a system should serve to generate the grammatical sentences of a language. Within early models of T-G grammar this total generating system was held to represent the "competence" of the native speaker of a language, and was contrasted with models which described only "performance," or the utilization of the internalized system to produce specific sentences at particular points in space and time.[52] In much recent literature the dichotomy between competence and performance has been shown to be less clear than originally supposed, and many aspects of language which were earlier claimed to be unsystematic are now claimed to be part of the broad linguistic competence of individuals.

The formal devices used within T-G theory for characterizing the relatedness among partially similar grammatical systems are intimately connected with the overall conception of linguistic systems proposed by the advocates of the model. As one particular integrated model of language structure, that presented by Noam Chomsky in *Aspects of the Theory of Syntax* (Chomsky, 1965), has been employed by a number of scholars who have dealt with dialect diversity, it is worth our while to examine it in some detail. A diagrammatic outline of this model is shown in Figure 22.

Grammatical systems of this sort can be roughly broken up into base, transformational, phonological, and semantic components. The base component consists of a set of phrase structure rules--context-free rules of the form A → B + C--where each symbol refers to a grammatical category standing for a class of lexical items or an individual lexical item. The effect of these rules is to substitute sequences of linguistic constituents for their superordinate category terms. The application of all of the phrase structure rules of a language is said to generate the set of "deep structures" of the language. These deep structures may themselves undergo one or more of a set of syntactic transformations which can delete, add, permute, or substitute elements in either deep structures or in structures produced by the application of other transformations. The members of the set of syntactic transformations are significantly ordered with reference to each other, and any number of them may apply in the derivation of a single output sentence.

The output of the set of syntactic transformations is a set of "surface structures"--structures whose formative elements are in the same linear order as in the output flow of speech--with each structure having assigned to it a structural description in the form of a branching tree diagram. The low-level constituents of these surface structures roughly correspond to morphemes and lexical items, each of which can in turn be represented as a linear sequence of phonological units (called variously phonemes, systematic phonemes, morphophonemes, etc.), where each unit represents a set of values for any of a finite

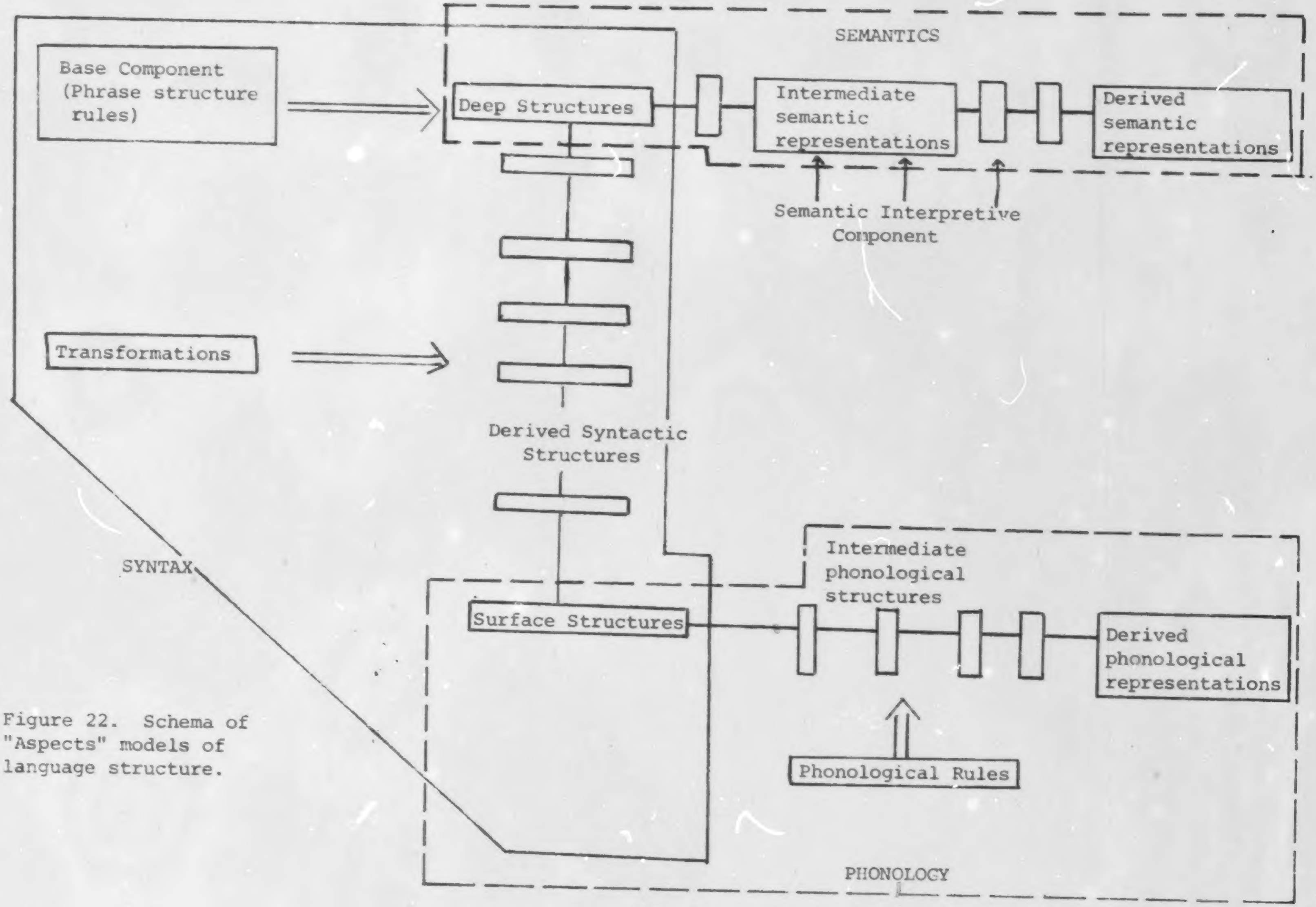

SEMANTICS
Base Component
(Phrase structure rules)
Deep Structures
Intermediate semantic representations
Derived semantic representations
Semantic Interpretive Component
Transformations
Derived Syntactic Structures
SYNTAX
Intermediate phonological structures
Surface Structures
Derived phonological representations
Phonological Rules
PHONOLOGY

Figure 22. Schema of "Aspects" models of language structure.

number of linguistically distinctive phonological variables, or "distinctive features." The linear sequences of these phonological units serve as the input to the phonological component, which contains another ordered set of context-sensitive rules to delete, add, permute, etc. specific features of individual phonological units. The application of these rules produces a set of derived phonological representations, the ultimate output of the phonological component of a generative grammar.

The semantic component of a generative grammar is by far its most controversial aspect, and many scholars have questioned the validity of postulating it at all.[53] Nevertheless, within the concept of grammar represented by Chomsky (1965), the set of deep structures produced by the operation of phrase structure rules is said to undergo semantic interpretation rules; the output of these is a set of so-called "derived semantic interpretations" which are intended to represent the semantic structures of the sentences of the language.

The model just described contains within it a number of bases for the comparison of partially similar grammatical systems. These systems can differ in the number and content of the units at a given level of analysis, as well as in the patterns of distribution of these units at that level. Systems can differ in the number and substance of the rules which relate strings of units at different levels of analysis, as well as in the conditioning factors which trigger the application of similar sets of rules. Moreover, sets of rules such as those which relate deep structures to surface structures can differ in the method in which they are applied (i.e. cyclically, simultaneously, or in a single non-repeatable ordered list), and, if applied cyclically, in the order of rules within a single cycle of application.

T-G grammar has made use of only a few of these potential areas of variability in discussing dialect diversity. Most transformational literature on dialectology has focused on the broadly phonological facets of linguistic diversity. Much of the literature has taken the approach of examining obviously related speech varieties, often geographically determined variants of a single "language," or independent "languages" which can easily be traced back to a common ancestor. T-G grammar attempts to characterize the differences in synchronic generative systems which can account for divergent phonological forms. Often such discussions are entered into not solely to clarify the description of dialectal situations, but also to defend any one of a number of theories concerning historical sound change.

The position which has emerged in a number of generative approaches to dialectology is that dialect differences can optimally be characterized in terms of slightly different stocks of phonological rules, and in different orders of application of sets of similar rules.[54] A number of sample applications of these principles are provided by Saporta, one such dealing with dialect differences arising through phonological merger in three dialects of Spanish.

	Castilian (C)	Latin American (LA)	South Chile (SC)
'Monday'	lúnes	lúnes	lúnes
'Mondays'	lúnes	lúnes	lúnes
'pencil'	lápiθ	lápis	lápis
'pencils'	lápiθes	lápises	lápis

Figure 23. Plural in three Spanish dialects
(from Saporta, 1965:219).

Within Saporta's framework two rules are needed to fully describe the facts given above:

Rule 1.

$$\text{pl} \rightarrow \begin{cases} \text{s} \ / \ \begin{Bmatrix} \text{V} \\ \text{é} \end{Bmatrix} \underline{\quad\quad} \\ \varnothing \ / \ \text{Vs} \ \underline{\quad\quad} \\ \text{es} \end{cases}$$

Rule 2.

$$\begin{Bmatrix} \theta \\ \text{s} \end{Bmatrix} \rightarrow [\text{s}]$$

The first of these rules states that the plural is manifested by the segment /s/ after unstressed vowels and stressed /é/, by Ø after a sequence of unstressed vowel followed by /s/, and as /es/ elsewhere. The second of the rules states that /θ/ and /s/ fall together and are realized as [s]. Saporta argues that dialect C is, in essence, the simplest of the three dialects, and the data from it can be explained purely in terms of Rule 1. Dialects LA and SC are more innovative than C, and require the application of both Rule 1 and Rule 2; however, they differ in the order in which the rules apply. Saporta claims that Rule 2 follows Rule 1 in LA, but precedes it in SC. In LA, then, the plural morpheme is originally realized the same as in C, but a subsequent rule (2) readjusts the output of Rule 1, reducing the number of possible surface segments. In dialect SC /s/ and /θ/ are merged prior to the application of Rule 1, thus reducing the number of possible segments which serve as input to Rule 1.

To many transformationalists then, dialects are considered to be related grammatical systems, sharing a substantial number of rules, and differing in either the total number of rules or the order in which these rules are applied. The relative "superficiality" of dialect differences within this model is captured by making the rules by which dialects differ late in order, and outside of the primary syntactic processes of the language. Dialect differences which can be accounted for within transformational grammar need not in theory be restricted to phonology, however. The same principles of rule ordering can pertain to syntactic processes. Thus if two linguistic systems share a common body of syntactic transformations presumably operating upon a shared set of underlying representations, and differ in the ordering of these transformations or in the presence or absence of a small group of transformations in addition to a commonly held core, they can be considered dialects of the same language.

One of the main drawbacks of the various attempts to account for dialect differences within transformational theory is the failure of the model to evaluate the relatedness of grammatical systems whose historical origins cannot be assumed. Like many other models we have discussed, T-G grammar starts by characterizing speech forms the relatedness of which is not open to question. For example, Saporta chooses for his examples data from three Spanish dialects, his goal being to characterize differences among what are, by common agreement, dialects of a single grammatical system. But what are the limits of a single grammatical system? Surely any speech forms, even if historically unrelated, will share some linguistic units or processes in common. At what point can we say that contrasted linguistic systems are sufficiently similar in terms of inventories of units, number and kind of linguistic processes, etc., so as to constitute unified systems? The model does not address itself to this question. What is claimed by transformationalists is that the grammatical systems which they describe are those of idealized speaker-hearers. But as William Labov has often pointed out, this position leads to the paradox that transformational grammar, in attempting to be a universal schema for representing language, is reduced to describing the structures of idiolects.[55] It makes sense within the model only to compare formal systems of idiolects and extrapolation to community based grammars is impossible.

T-G grammar is as rigid as many earlier models of linguistic description in presupposing a fully homogeneous linguistic code as the primary object of linguistic investigation. If the possibility of the existence of dialects is admitted, each of these dialects must have a regular structure and sentences produced within them must be explainable in reference to a fixed body of highly structured rules. Systematic variability exists only between idiolects, and is not part of the "competence" of any single individual. Factors of language use not explainable in terms of a regular set of internalized rules must be explained in terms of the vagaries of linguistic "performance." Social factors which can contribute to the selection of alternates with regard to grammatical variables are not considered to exist at all. Transformational generative grammar gives form then to the dichotomy between language and dialect only by comparing autonomous grammatical idiolects, and by subsuming similar sets of such idiolects into larger order classes. The theory does not describe either how this grouping of idiolects is to be carried out or how similarity between grammatical systems can be evaluated. The model has in fact confined itself to illustrations of dialect differences which skirt the definitional problem.

2.5. Sociolinguistic Models of Language and Dialect

If there is any one trait which has been held in common by most twentieth century schools of linguistic description, it has been the adherence to the tenet that a language or dialect is a cohesive entity which may be described without reference to entities external to it. Language is, to be sure, an inherently social entity, and this fact has been denied by few linguists,[56] but the description of its use in the world of social interaction has been held to be beyond the domain of linguistic investigation. While minority claims to the futility of such a vigorously "antisocial" model of language have been persistently voiced since at least the end of the 19th century,[57] it is only in the past decade that formal models have been developed which attempt to integrate the formal structures of linguistic systems with the social orders in which these systems are employed. The attempt to view linguistic systems in the context of the social settings and social functions in which they are used can be considered to constitute a "sociolinguistic" approach to language study. There are any number of such approaches, differing among themselves in the formal devices suggested for capturing sociolinguistic correlations, the particular types of covariance held to be of interest, and the overall models of what constitute "languages" designed in response to an enlarged conception of linguistic behavior.

On the most simplistic level, the point of departure of recent sociolinguistic investigation has been the observation that within communities linguistic usage is seldom fully homogeneous. Geographical differentiation of language has, of course, been well known and recorded in dialect atlases for some time.[58] But the observation that linguistic usage within a single community varies along a number of discrete dimensions, and that the speech of even single individuals may vary with social context, has only recently been rendered theoretically describable by means of mainstream linguistic methodology. Such variation had previously been observed, but was frequently asserted to be marginal to the central linguistic "facts" of a language. If it was at all acknowledged it was by asserting that the linguistic usage of an individual or community may encompass multiple norms, but each of these norms should be considered a cohesive uniform entity. Data incompatible with such "multi-dialectal" norms were considered deviant, substandard, or not to exist.[59] Not infrequently such data were labeled as being in the domains of "stylistics," "performance," or something of the sort, and removed from linguistic consideration. In contrast to such views, some current approaches to language assume that linguistic heterogeneity is an integral part of linguistic competence, and that the ability to use language involves the knowledge to systematically manipulate areas of linguistic variability in response to complex social environments.

In essence, modern sociolinguistics asserts that linguistic performance is sufficiently complex to preclude its being described purely in terms of the structure of an isolatable linguistic code. It involves the ability to select and manipulate options built into formal linguistic devices, with the choice among these options reflecting the social backgrounds, circumstances, aspirations, and attitudes of the speakers and the people with whom they come into contact.[60]

The incorporation of what we can call "systematic variability" into linguistic description is a radical departure from much of earlier linguistic practice. Within transformational grammar, for instance, fundamental processes of a language (i.e. transformations) are held either to apply or not apply. A linguistic structure may meet the structural requirements for the application of some rule, or it may not. This is an all or nothing situation. The rule may be written so as to appear "optional", but in that case it becomes a matter of accident as to whether the option is actually employed. Until quite recently transformational rules were not formulated in a way which allowed the statement of factors which can influence positively or negatively) their operation. This difficulty in capturing "variability" is not limited to transformational grammar. Grammars characteristically operate in binary terms, with continua as to the operations of rules, allowable only with great difficulty. Bloomfieldian, neo-Bloomfieldian, and Transformational linguistics have as a rule allowed for the description of variation in linguistic phenomena which can be predicted through internal linguistic considerations. They have not, however, accepted explanations for such variation in which linguistic phenomena systematically covary with social phenomena, and in which the observer is commonly presented with continua in the frequency of occurrence of the options of the linguistic variable. Linguists are most comfortable with categorical statements, whereas sociolinguistic generalization invariably needs to make recourse to frequencies of occurence.

By way of illustrating the sort of variable phenomena which we are referring to, we offer examples from Indian languages. Hindi has an inventory of consonants which have been historically derived from the consonant system of Old Indo-Aryan. The pronunciation of many speakers includes a number of consonants which are not part of the indigenous systems, and which have been introduced into the language through the absorption of Persian and Arabic loan words. The presence or absence of these consonants is seldom categorical within a community, and tends to be correlated with the degree of education, sex, and social background of the speaker. Thus many Hindi speakers in certain contexts employ the phonological elements *f, x, ɣ,* and *z* in the place of *ph, kh, g,* and *j* respectively (e.g. *fir* ~ *phir* "again, then;" *xudā* ~ *khudā* "God;" *baɣair* ~ *bagair* "without;" and *bāzār* ~ *bājār* "market"). Not all of these alternations are of equal currency, and among the above examples the pronunciation *bāzār* is far more common than *bājār*. Moreover, the amount of use the competing consonants in an alternation receive is likely to vary with the degree of formality of the speech event. Thus any of a number of factors influences the *likelihood* of the utilization of a Perso-Arabic phonological unit in the place of a corresponding indigenous Indo-Aryan one.

The description of phonological alternations such as these has posed problems for linguists in the absence of a linguistic description which takes into account the conditioning factors influencing the selection of alternate phonological units, Linguists have been forced to consider the situation to be one of a language having multiple, albeit overlapping, phonemic inventories, and to assume that if a speaker has the ability to switch between the alternates, he is, in effect, switching codes. Such a speaker is considered multidialectal, even though the dialects are similar except for the phonemic inventories. The linguistic competence of such a person parallels, on a smaller scale, that of an individual who is competent in both French and Spanish.

A different kind of sociolinguistic variability is exemplified by relative clauses in Hindi. It is characteristic of Hindi relative clauses that the actual relative pronoun enjoy great freedom of position within the relative clause, and that it not be restricted to clause initial position.[61] Moreover, sentences are possible both in which the relative clause precedes the main clauses, and in which the opposite is the case (e.g. *mãī ek ādmī ko jāntā hũ jo purāne bāzār ke pās rahtā hai* "I know a man who lives near the old market" ~ *jo ādmī purāne bāzār ke pās rahtā hai use mãī jāntā hū̃)*. In addition to these structures, Hindi also employs a type of relativization in which the noun which is shared by the main and dependent clauses is first relativized, made the subject of a copulative predication, and lastly questioned (e.g. *purāne bāzār ke pās jo ādmī hai na, use mãī jāntā hū̃* "near the market which man is, right?, him I know = I know the man who lives near the old market). This last construction, with its seemingly empty phrase *jo hai na,* is widespread in both non-standard varieties of Hindi and in spoken varieties of educated speakers of the language. It is, we suspect, statistically one of the most common of relative constructions, in at least certain styles of the language. Yet the construction is virtually absent in other more formal styles of the same speakers. The constraints on the usage of the construction seem to involve the socio-economic background of the speaker in some cases, and the formality of the context for other speakers. Among no speakers would the construction be used in formal written Hindi, regardless of the degree to which it might be employed in spoken language. Once again we are presented with a readily describable set of linguistic phenomena, but one in which it is necessary to make recourse to social considerations to explain the distribution of what appear to be synonymous expressions. Within transformational grammar, which has little difficulty in formulating rules to explain the syntactic relationship between the competing structures,[62] there is no obvious way of building into the structural description for the rule a set of social conditioning factors.

One of the major contributions of recent sociolinguistic theory has been the development of techniques which can be used in describing socially motivated linguistic phenomena. A major advance has been the advocacy of the notion that certain linguistic processes are inherently variable, and that some environments increase or decrease the likelihood of the occurrence of the process.[63] Once this has been admitted then it is theoretically feasible to describe quantitatively the extent to which given social factors influence the application of the rules.

Theoretically any potential aspect of linguistic variation can admit of such treatment--a range of phonetic realizations of a single lexical item, the employment of different surface case markings (e.g. dative versus genitive) for the expression of an abstract syntactic function (e.g. the subject of verbs of perception), or a movement transformation shifting the order of some linguistic constituents. Labov, in his important work, *The Social Stratification of English in New York City*, notes many examples of phonological variability: the presence or absence of final or preconsonantal /r/ in words such as *car* and *card;* the height of the vowel in such words as *bad, bag, ask, pass, cash,* and *dance* (ranging as high as [I^ə] and as low as [a:]); the height of the vowel in *caught, talk, awed, dog, off, lost,* and *all* (ranging from [ʊ^ə] to [ɑ]); the pronunciation of the initial consonants in *thing* and *then* (ranging from [t̪] to [θ] in the first case and [d] to [ð] in the second). On the morphological level, variation can be seen in the alternation between the postpositions *ke liye* and *ke vāste* in Hindi (e.g. *āp ke liye* ~ *āp ke vāste* "for you") or the similar alternation of *nī māṭe* with *māṭe* in Gujarati (*jovā nī māṭe* - *jovā māṭe* "in order to see"). The optional marking of *cāhiye* "is/are required/ necessary" in order to agree with subject nouns in number in Hindi (e.g. *mujhe pustak cāhiye* - *mujhe pustakẽ cāhiyẽ* "I need/want a book" - "I need/ want some/the books.") is yet another example of morphological variability. On a syntactic level the extrapositioning in Hindi of sentential constituents to a position after the finite verb of a sentence (e.g. *mãĩ ne kām kiyā* - *kām kiyā mãĩ ne* "I did the work") exemplifies linguistic phenomena which admit of socially conditioned variation.

As already noted, phenomena of this sort while being noticed in the linguistic literature, have generally been considered to be random (that is, the alternates are in free variation), the result of deficient knowledge of the language, or governed by extra-linguistic factors. It has traditionally been acceptable for linguists to describe variation which is conditioned by other linguistic phenomena. For example, it is common practice to describe a phonological alternation between *t* and θ if it can be shown to be a function of the position in the word which these sounds occupy, the nature of the preceding or following sound, or the grammatical category of the word or morpheme in which the sound occurs. But it has not, on the other hand, been common practice to describe linguistic variation whose conditioning factors are "non-linguistic." Correlations of linguistic variation and independently motivated social variables have been shown to be linguistically productive. Labov, for instance, has demonstrated that there is a correlation between the frequency of appearance of postvocalic *r* and socio-economic class in New York City. The higher the socio-economic class, the higher the percentage of times in which the *r* appears. Moreover, the occurrence of *r* is correlated with the contextual style in which the sample of the informant's speech is elicited. The more formal the style, the less likely the informant is to use the *r*-less variant. The correlation between percentage of times in which *r* is realized postvocalically and socio-economic class and contextual style is shown below in Figure 24.

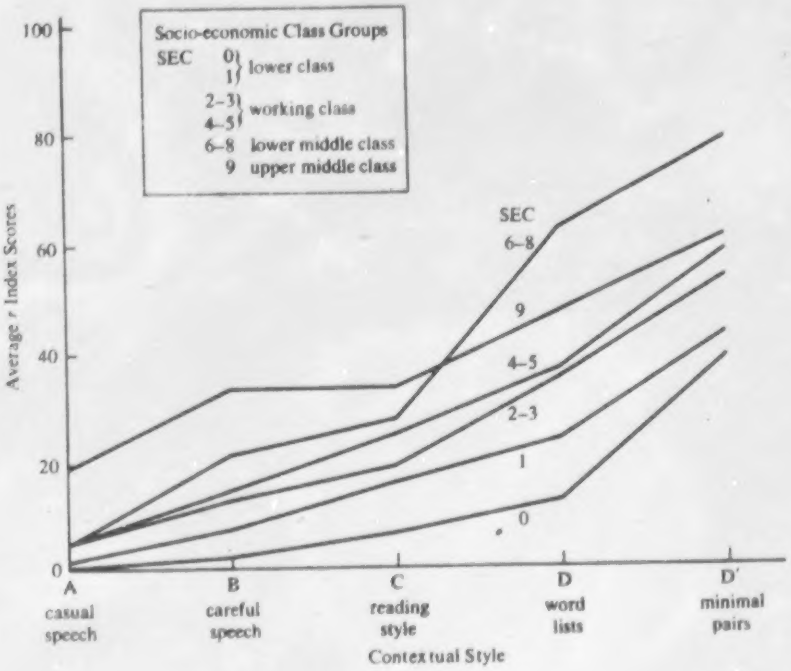


Figure 24. Class Stratification diagram for *r*. From William Labov, "Phonological Correlates of Social Stratification," in *The Ethnography of Communication, American Anthropologist* 66, no. 6. Part 2, p. 171. Cited by Wolfram and Fasold (1974:87).

In another important work Labov has demonstrated a different type of "inherent" variability with regard to contraction and deletion of the English copula. He is able to show that "wherever S[tandard] E[nglish] can contract, B[lack] E[nglish] V[ernacular] can delete *is* and *are*, and vice versa; wherever SE cannot contract, BEV cannot delete *is* and *are*, and vice versa."[64] Another common, although caricatured, variable in English is the use of *ain't* which, according to Wolfram and Fasold,[65] occurs in virtually every nonstandard American English dialect. The list of socially conditioned areas of inherent variability in English alone is quite large, and a description of many of them is provided in Wolfram and Fasold.[66] There is no reason to assume that a similar list cannot be constructed for all languages of the world. We will further discuss such areas of variability in South Asian languages in Chapter 5.

Once areas of "inherent" linguistic variability have been isolated, the next task of sociolinguistics is the demonstration and documentation of factors which increase or decrease the likelihood of occurence of one or more alternates of the variable. Characteristically, both linguistic and social "constraints" play a role in determining the alternate of a linguistic variable which is to be used in a specific context. Wolfram has shown, for instance, that the presence or absence of final consonant cluster simplification in Detroit black speech is influenced by the social class of the speaker, whether the initial sound of the following word is a vowel or consonant, and whether the final consonant to be potentially deleted is a realization of the past tense morpheme $\{-ed\}$.[67]

Moreover, more than one social variable may influence a given linguistic variable. In the data given in Figure 24 it is the *interaction* between contextual style and socioeconomic class which ultimately is correlated with the specific percentage of realization of *r* by a given speaker. The interaction between numerous variables can be quite complex. First of all, the social variables characteristically are of unequal strength. It is then meaningful to speak of a hierarchically arranged set of constraints which covary with purely linguistic phenomena. It is the contention of some linguists that it is even possible to quantify the degree to which any given constraint effects the likelihood of occurrence of an alternate of a linguistic variable. When numerous social variables interact with linguistic ones, each alternate of the social variables can be assigned a probability which it contributes to the likelihood of occurrence of the linguistic

	Social Classes			
ENVIRONMENTS	UPPER MIDDLE	LOWER MIDDLE	UPPER WORKING	LOWER WORKING
Following vowel Final member is *-ed*	.07	.13	.24	.34
Following vowel Final member is not *-ed*	.28	.43	.65	.72
Following consonant Final member is *-ed*	.49	.62	.73	.76
Following vowel Final member is not *-ed*	.79	.87	.94	.97

Figure 25. Linguistic effects on frequency of final consonant cluster simplification in Detroit black speech (adapted from Walt Wolfram, *A Sociolinguistic Description of Detroit Negro Speech*. Washington, D. C.: Center for Applied Linguistics, 1969: 59-69. Cited in Wolfram and Fasold, 1974:132.

alternates. The ultimate probability of occurrence of any alternate of the linguistic variable in a particular context is a function of the composite probabilities of the individual constraints. A sample ordering of three constraints on the operation of a variable rule is given in Figure 26.

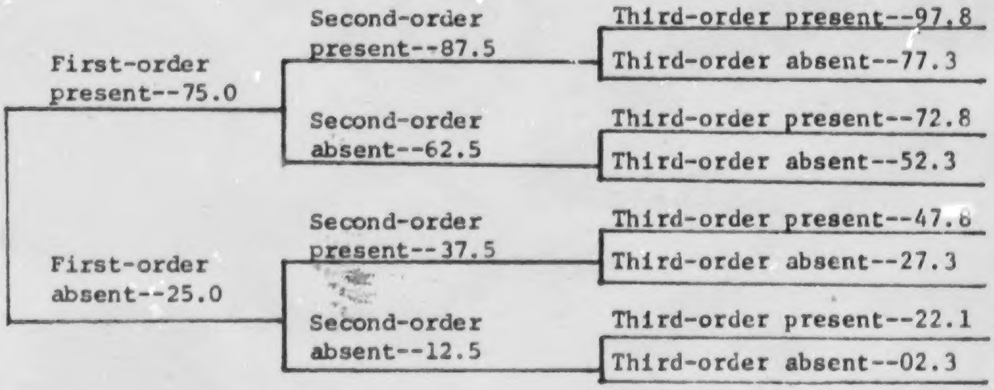

First-order present--75.0
Second-order present--87.5
Third-order present--97.8
Third-order absent--77.3
Second-order absent--62.5
Third-order present--72.8
Third-order absent--52.3
First-order absent--25.0
Second-order present--37.5
Third-order present--47.8
Third-order absent--27.3
Second-order absent--12.5
Third-order present--22.1
Third-order absent--02.3

Figure 26. Hypothetical ordering of three constraints on a variable rule. From Wolfram and Fasold (1974:108).

This chart could, hypothetically, represent the frequency of deletions of the final member of terminal consonant clusters in English. The first-order constraint might indicate low degree of education (as opposed to non-low-education), the second rural (as opposed to urban), and the third male (vs. female). In such a case Figure 26 would indicate that poorly educated rural males delete final consonants in clusters 97.8% of the time, whereas educated urban females do so only 2.3% of the time. The same chart also indicates that the sex of the speaker plays less of a role in determining the frequency of deletion than does the feature rural vs. urban, which in turn is of less significance than the education of the speaker. Moreover, by using techniques developed by Cedergren and Sankoff[68] it is possible to calculate the specific probability of deletion contributed by each of the three constraints (.89 for the first, .57 for the second, and .16 for the third.)

There is no fixed inventory of social variables which potentially can interact with linguistic variation, although there are a number of them which most readily spring to mind--socio-economic background, amount of education, ethnic background, age, generation, sex, urbanization, degree of awareness of the speech event on the part of the speaker, audience of speaker, etc. Other social variables are also possible, and the ones which have been cited here are merely those which have proven most useful in post-sociolinguistic investigation in the west. There is no reason to believe that there exist universally valid social categories with respect to their utilization in conditioning linguistic processes. It is reasonable to surmise that linguistic variation in non-Western societies may in part be conditioned by social variables having little or no predictive value in the west. The determination of social variables which are most suitable for use in describing the linguistic usage of a community must be done in relation to that community, and not involve the superimposition of a priori categories. The ultimate goal of sociolinguistic investigation is, after all, the ability to make highly predictive generalizations about some aspects of potential utterances on the basis of the social context in which the utterance is spoken. It is only common sense that the social factors which have the highest predictive value are likely to be those possessing psychological value for the speakers in question. The superimposition of unmotivated or culturally alien variables is likely to yield "explanations" of linguistic variation which are less than optimal. The opposite caution is, however, equally in order. The utility of a social variable in a non-linguistic context, regardless of its utility in explaining societal aspects of a community, does not insure that the variable is useful in describing sociolinguistic covariation. On the contrary, the mechanical employment of such a variable leads to an obscuring of patterns of covariance which are much more deeply rooted in the community.

In dealing with patterns of covariance of the sort described here, it is important to be cognizant of the distinction between empirically valid patterns of variation in observable linguistic behavior, and patterns of variation in the subjective attitudes of speakers toward his variation. Patterns of sociolinguistic variation which are deeply engrained in a community may not be fully perceived by the individuals in that community. The linguistic behavior of an individual is, in part, a function of a complex set of social, environmental, and contextual constraints, and the perceptions of an individual concerning those constraints may be less than fully accurate. An individual's perception of the linguistic variation around him often is concentrated on those features which are accorded some sort of social value, be it positive in the case of socially prestigious items or negative in the case of stigmatized ones. It may fully pass the individual's attention that precisely those items which he most stigmatizes are used in his own speech a high percentage of the time. It is for this reason that Labov has distinguished among three socially diagnostic types of linguistic features which are referred to as *social indicators, social markers,* and *social stereotypes.*[69] The first of these terms refers to a linguistic feature whose use is diagnostic of some aspect of the speaker, and whose social value features may be unnoticed by him. The second term refers to a linguistic feature which has social value attached to it and whose use may affect the listener's perception of the social qualities of the speaker. The last of these terms refers to a linguistic feature which has come to be generally considered a characteristic form of verbal behavior of individuals having particular social properties. It is not necessarily the case, however, that the class of individuals said to be characterized by the use of a social stereotype will, in fact, use that linguistic feature more than will other social groups.

Linguistic features which have become social stereotypes, or in some cases, social markers, are important in sociolinguistic variation. It is these items which often bear the brunt of the speaker's attempt to manipulate their (and other's) language in particular contexts. If an individual perceives that a stigmatized linguistic feature identifies him as having undesirable social characteristics, he may consciously attempt to eliminate that feature from his speech. In such cases his success in this attempt **is** likely to increase with the formality of the speech event, since in formal circumstances the degree of awareness of his own speech is greatest. Social indicators, on the other hand, which are unnoticed by their speakers, may be more evenly used than social markers through a wide variety of stylistic contexts by their speakers.

Sociolinguistic approaches to linguistic description thus need to take into account a much larger range of linguistic phenomena than other sorts of approaches have to. They need not only describe observable patterns of variation in purely linguistic phenomena, but attempt to place these patterns in a social context. This social context incorporates aspects of the speaker's and listener's backgrounds, the immediate environment in which the speech act occurs, the speaker's awareness of the speech event (i.e. the stylistic context), as well as the speaker's attitudes about his own and other's speech. The adoption of such approaches stretches considerably traditional concepts of what is meant by the terms "language" and "dialect." No longer can a speaker be said to control a single "dialect" or "idiolect." Rather, he controls an array of styles and a set of criteria by which he can switch among them. A community will not be characterized as having a single homogeneous speech variety, but will possess a continua of varieties, with individuals in that community controlling different segments of the continuum. The composite linguistic competence of a community would then have to include a specification of the constraints which govern the selection of linguistic variants by members of the community. These constraints will necessarily affect different members of the community to varying degrees. Under such circumstances a minimal linguistic variety becomes not a dialect or idiolect, but rather a "lect", namely that speech variety used by an individual in a particular set of social circumstances.

The total repertoire of a social group becomes not the single cohesive entity which we think of when we use the term "language," but rather a body of linguistic traits shared by the total number of "lects" and a set of variable linguistic traits whose use is distributed within the community. The writing of "pan-lectal" grammars, a field recently popularized by C. J. Bailey, is a task of far greater complexity than the description of static linguistic codes, and one which has only recently begun to be seriously attempted.

2.6. *The Linguistic Basis of Typological Studies of South Asian Languages*

In this section we would like to turn our attention to some fundamental notions in the classification of languages in South Asia. We attempt to discuss the theoretical frameworks within which much of the early classification of South Asian languages and dialects was carried out, and to point out ways in which these frameworks were inadequate for describing certain aspects of language use in the area. This will serve as an introduction for the more detailed examination of language classification in South Asia to follow in Chapter 3.

There has been in the past no paucity of typological studies of South Asian languages. Within western scholarship attempts to fix the genealogical places of various South Asian languages have gone on as long as there have been descriptions of South Asian languages. The earliest of these grew out of attempts to establish the place of Sanskrit in relation to the known classical and vernacular languages of Europe. Sir William Jones, writing in 1786 and drawing upon over a century of rather unsystematic observations on South Asian languages by Jesuit missionaries, is considered to have initiated the scientific comparison of these languages with his oft quoted statement that Sanskrit in relation to Greek and Latin "bears a stronger affinity, both in the roots of verbs and in the forms of grammar, than could have been produced by accident; so strong, indeed, that no philologer could examine all three without believing them to have sprung from some common source, which perhaps, no longer exists; there is a similar reason, though not so forcible, for supposing that both the Gothick and the Celtick, though blended with a very different idiom, had the same origin with the Sanskrit."[70] Other European scholars, most notably Friedrich von Schlegel, Franz Bopp, A. F. Pott, and T. Benfey, expanded upon Jones' observations, and provided the detailed basis for the comparative study of "Indo-European," a study which required intimate knowledge of Sanskrit. It was thus Sanskrit among Indian languages which was first known in the west, and the typology of South Asian languages grows out of the attempt to define the position of Sanskrit in relation to other languages such as Latin, Greek, etc. The study of vernacular South Asian languages came later. In the case of Indo-Aryan vernaculars, comparative word lists were drawn up and grammatical sketches published by missionaries in the seventeenth and early eighteenth centuries. These became more widespread in the middle of the nineteenth century, and flowered at the end of the century with the publication of grammars and dictionaries of a large number of standardized Indo-Aryan languages. These grammars served as the stepping stone for the "comparative" analysis of the contemporary Indo-Aryan vernaculars, the results of which are consolidated in such important works as Kellogg, *A Grammar of the Hindi Language*, 1875[71] (which is essentially a comparative grammar of all of the Indo-Aryan vernaculars east of the Gujarat and west of Bengal), John Beames *A Comparative Grammar of the Modern Aryan Languages of India: To Wit, Hindi, Panjabi, Gujarati, Marathi, Oriya and Bengali*, 1872-79,[72] and Sir R. Turner, *A Comparative Dictionary of the Indo-Aryan Language*, 1966.[73] Much of the early work on the comparative typology of Indo-Aryan was summarized and expanded in the introductory volume of the *Linguistic Survey of India* (1903-28)[74] and in Grierson's monograph, "On the Modern Indo-Aryan Vernaculars."[75] Parallel work on Iranian typology has also been carried out during the past two hundred years, beginning as a sporadic series of linguistic accounts of the language of the Avesta during the middle of the 18th century (the most notable of which was by du Perron in 1771),[76] and expanding with the development of comparative Indo-European grammar (early major contributions were here made by Sir William Jones, Rasmus Rask, and Niels Ludvig Westergaard). Major work in the development of comparative Iranian linguistics followed with the publication of such grammars as A. V. W. Jackson's work on Avestan (1892),[77] Reichelt's *Awestisches Elemantarbuch* (1909)[78] and Christian Bartholomae's *Altiranisches Wörterbuch* (1904).[79] During the 19th century numerous grammars were written on middle and new Iranian languages and a major synthesis of these works achieved in Wilhelm Geiger and Ernst Kuhn's *Grundriss der Iranischen Philologie.*[80]

As in the case of Indo-Aryan, speculation as to the internal groupings of the Dravidian languages dates back virtually as far as the existence of grammars of each of the main members of the family. In 1816 Ellis wrote an introduction to A. D. Campbell's Telugu grammar[81] in which he compares lexical items in Tamil, Kannada, and Telugu in order to demonstrate a linguistic affinity among the three languages. Stevenson (1852)[82] sought to show that these and other similarities could not be explained on the basis of a corruption of Indo-Aryan material. Caldwell in 1856[83] attempted the first systematic genetic classification of the Dravidian languages, a classification which later proved to be the basis of Grierson's classification of the Dravidian languages in *Linguistic Survey of India*. Since Grierson's time a substantial amount of grammars and linguistic studies of individual Dravidian languages have been written, and major syntheses of their implications for the comparative structure of the Dravidian languages achieved in Burrow and Emeneau, *A Dravidian Etymological Dictionary* (1969),[84] Bh. Krishnamurti's article, "Comparative Dravidian Studies",[85] and Zvelebil's *Comparative Dravidian Phonology.*[86]

Comparative typological studies of comparative and typological structure of the Munda and Tibeto-Burman languages of South Asia are considerably sparser than materials for Indo-Aryan, Iranian, and Dravidian. Work in this area has been hampered by both the absence of long literary traditions for most of the languages, by the comparative inaccessibility of many of the groups speaking the languages, and by the absence of standardization movements for many of the speech forms. Most of the early work on the Munda languages, dating back approximately 150 years consists of word lists and more extensive grammatical descriptions. The first significant classification of the Munda languages was accomplished in the *Linguistic Survey of India* (volume 4, edited by S. Konow, Calcutta, 1906). Later work on the language classification of the Munda languages was summarized in H. Maspero's "Les langues mounda," in Meillet and Cohen, eds., *Les langues du monde* (Paris, 1952). Important later work on the classification of the Munda languages includes Pinnow's *Versuch einer historischen Lautlehre der Kharia-Sprache* (1959)[87] and a number of papers by Norman Zide.[88] Theories on the external relations of the Munda languages have been presented by Pinnow[89] and F. B. J. Kuiper.[90] Material on the linguistic classification of the Tibeto-Burman languages of South Asia is also extremely limited. The vast majority of known Tibeto-Burman languages are represented in extremely small corpora of data, and speculations on the internal relations of these speech forms is often unwise. According to Roy Andrew Miller, in his survey article on linguistic research on the Tibeto-Burman languages,[91] the major task of language classification of the area has only been seriously begun with the work of Robert Schaefer in the 1950's and 1960's,[92] as at the time of the *Linguistic Survey of India* insufficient data was available to make scientific typology possible. Schaefer's important pioneering work has been followed by, according to Miller, Uray Géza,[93] Róna-Tas Andras,[94] and Nishida Tatsuo.[95] Much of Schaefer's work over the past 25 years has been gathered together in his Introduction to *Sino-Tibetan* (Wiesbaden, 1966).

It seems to us not unreasonable to assert that much of the development of orthodoxy for South Asian language classification (with the exception of Tibeto-Burman) took place in exactly the period of time which witnessed the development of both the so-called comparative method of linguistic reconstruction and the growth of Indo-European studies. In fact it is well known that the acquisition of detailed knowledge of early members of what later came to be referred to as the Indo-Iranian branch of the Indo-European family of languages, provided a major impetus for the development of the Indo-European studies. Indological studies in the nineteenth century stood at the vanguard of Europe's rapid expansion of its linguistic horizons to the east and the development of a linguistic methodology capable of structuring the obvious resemblances between the known ancient European languages and the newly "discovered" Asiatic languages into a coherent whole. The early model chosen to achieve this structuring was the so-called Stammbaum or branching-tree theory of linguistic history.

As we have already stated, the period of time in which the first major classifications of South Asian languages were carried out is essentially the early period of comparative Indo-European languages. Grierson, and it is, after all, Grierson from whom most of our classifications stem, makes it quite clear in the introduction to *Linguistic Survey of India*, that he made extensive use of all of the available grammars, both of single languages, and of related languages in arriving at the general outlines of his taxonomies. In the case of Indo-Aryan for instance, Grierson, referring to Beame's *Comparative*

Grammar of the Aryan Languages and Hoernle's *Grammar of Eastern Hindi Compared with the Other Gaudian Languages*, states that "these two excellent works, each a masterpiece in its own way, have since been the twin foundation of all researches into the origin and mutual relationship of the languages of the Indo-Aryan family of speech."[96] The framework of classifications in Beames, Hoernle and other similar works is a Stammbaum model, in which the terminal nodes represent the standardized vernaculars of North India. It seems to us that one of the implicit aims of the *Linguistic Survey of India* was to extend the Stammbaum model to encompass a wide range of non-standardized vernaculars, to firm up those branches of assorted Stammbaums which could not be established conclusively on the basis of old information. Presumably this implied being able to establish a vertical hierarchy of language varieties, so that any regional vernacular could be included under the domain of a superordinate, and presumably standardized, language. The opposition between language and dialect was in part intended to correspond to relative levels of height within this hierarchy, although here Grierson points out numerous practical problems in precisely locating the point of division between the two terms and in using these terms to characterize specific speech forms. In expanding the range of traditional classifications of South Asian languages to include spoken vernaculars, Grierson makes use of techniques which predate the emergence of modern dialect geography, and which would yield Grierson the maximum amount of usable material in a reasonable period of time. Grierson clearly was seeking to make his classifications on the basis of cross-dialectal phonological and morphological correspondences. He therefore selected his material and methods of elicitation in order to give him as wide a basis for comparison as possible, and to enable him to make decisions about the classification of languages and dialects within a Stammbaum framework.

We do not wish to give the impression that the entire history of classification of South Asian languages has been passed down to us unaltered from Grierson. We do claim, however, that Grierson's classifications have in a sense served as the starting point in South Asian typology, and that the general framework, including views on what it means for languages to be related, on what is understood by the opposition of language and dialect, and in many cases the terminology itself which is used to designate classes of languages (e.g. the term Bihari, to cover a number of Indo-Aryan language varieties in Bihar and Eastern Uttar Pradesh) stems from Grierson. Moreover, many of the strengths and weaknesses of work in language typology stem from the strengths and weaknesses of Grierson's framework.

All theories of language start with certain assumptions and axioms, and proceed to make predictions about the structure of real-life language forms. For any such theory there is a body of data which is readily describable and a body of data which is not. The Stammbaum model of linguistic history, like any other model of language change has a core of data which is central to the theory, and whose describability is cited to validate the entire theory. Within Indo-European grammar this core of information has tended to be the inventories of the phonological units within a dialect, the paradigmatic relations holding between these units, inventories and paradigmatic relations among inflectional morphemes, and the presence of a common core of lexical items. This last case generally has been restricted to only a subset of core lexical items within linguistic systems. The model has sought to focus on precisely those aspects of languages which can be cited to demonstrate an historical continuity and relatedness among language varieties. The successful cases of such an approach have been enormous. Historical linguists can point with great pride to a wealth of reconstructions of protoforms in a large family of languages, Indo-European being the most widely known, and cite seemingly regular sound laws relating this protoform to a wide variety of contemporary forms in modern vernacular and standard languages through any of a large number of long and often tortuous routes. By postulating these rules, linguists were able to make explicit predictions about the actual forms in a group of languages, and prodded to revise the rules wherever their predictions could be shown in disagreement with the known facts. In short, the adoption of such a model has had its greatest success in demonstrating historical continuity among languages, and in explaining the differences between related languages in terms of "exceptionless historical rules." There are, however, numerous widely documented aspects of language use in South Asia which cannot be accounted for within the model. What follows is a description of some of the more important of these.

1. *Inability to handle areal features.* It is well known that in many parts of the world adjacent languages of diverse historical origin will often show a mutual convergence of linguistic features. Such an area (or *Sprachbund*) can be characterized as having shared lexical items, phonological, morphological, syntactic properties, or in fact share any isolated features of linguistic systems. The existence of one major *Sprachbund* encompassing most of South Asia has been extensively described by Bloch, Emeneau, Kuipers, Barannikov and others,[97] and can be characterized by a) a common core of phonological units (e.g. retroflexion), b) a shared set of syntactic constructions (e.g. the conjunctive/absolutive forms, the use of inchoative constructions,) c) shared types of lexical items--onomatopoetic terms, echo words. It is not unreasonable to suspect that other smaller range *Sprachbunde* are also postulatable at areas of convergence between groups of diverse historical origin, i.e., Himalayan *Sprachbund,* Northwest frontier, Indian-Burman border, Srilanka, Pakistan-Afghanistan frontier. It is also the case that linguistic features are often present in language forms which are not juxtaposed in a strict geographical sense. This is often the case when diverse language varieties borrow features from a superordinate prestige language. Such is the case both with regard to large numbers of Perso-Arabic loan words, and with them, Perso-Arabic phonological and morphological units in an area extending from North Central Africa to Southeast Asia. A similar case is in the independent borrowing of Sanskrit tatsama items in educated styles of a large number of South Asian languages as diverse as Pashto, Assamese, Nepali, and Telugu.

2. *Substratum of village dialects.* In large portions of South Asia is totally impossible to construct a model of standardized languages directly subsuming discrete classes of regional dialects. Most of the spoken vernaculars of North India, for instance, comprise a virtual continuum of speech forms extending across the subcontinent from Bombay or Ahmadabad in the west to Calcutta in the east. On a "grass-roots" level, adjacent links in the chain differ by only small scale linguistic features. Such small scale differences have little effect in influencing inter-group communication between adjacent links on the chain, but become compounded with greater distance. One is thus dealing with a diffusion of interconnected speech varieties among whom mutual

intelligibility is at least in some sense a function of distance. Among the adjacent dialects of this chain one can observe a great deal of structural similarity in phonology, morphology, syntax and vocabulary. The Stammbaum model predicts that the greatest divergence between language varieties should be observed horizontally at a given level cutting across branches of the tree. The North Indian language continuum, however, is characterized by blurring the horizontal boundaries across language groupings. Moreover, the Stammbaum model has few or no means at its disposal for describing the horizontal spread of linguistic features which clearly must operate in the development of such continua.

3. *Stylistic variation within dialects or idiolects.* Virtually all known standardized South Asian languages have sharply distinguished styles or registers. These distinct styles are often characterized by differences in phonology (as for example the pronunciation vs. non-pronunciation of Sanskritic loan borrowings in Gujarati as reported by Pandit),[98] morphology (the selection of *ke vāste* vs. *ke liye* in Hindi, syntactic (i.e. use of *jo hai* in colloquial Hindi-Urdu vs. lack of it in educated) and lexicon (i.e. use of classical Tamil forms and expulsion of obvious Sanskritisms in literary Tamil vs. higher percentage of loans in colloquial.) These multiple styles will often be controlled by different speakers within a community, with the conditioning factors social (i.e. caste dialects, Brahmin vs. non-Brahmin/male - female) or spoken by the same individuals in a socially determined set of circumstances (i.e. formal vs. non-formal.) In some cases the stylistic versions of what can be considered one language will be sufficiently great as to produce non-intelligibility. Cases which come to mind here include colloquial vs. literary Tamil, and the three-way distinction of colloquial *khaRī bolī*, *sāf* Urdu and *shuddh* Urdu, with virtual noncomprehensibility often holding between extreme versions of Hindi and Urdu. Often the varieties of language spoken by a group at a given social level in one "language" will have greater structural affinities with other dialects of a parallel social group within another language than with other social or stylistic registers within the same language.

4. *Codeswitching.* Largely through the pioneering studies of John Gumperz during the 1950's and 1960's[99] we have information of South Asian speech patterns where individuals command as part of their

linguistic repertoire the ability to switch back and forth, often within the same utterance, between what otherwise needs be considered autonomous codes. Often such language use is sufficiently heterogeneous as to preclude being labelled in any one code or other. Such circumstances often occur in areas of language convergence both within a single language family (as in Hindi-Punjabi) or between numerous genetic stocks (Kannada, Urdu, Marathi). Such code switching occurs with individuals demonstrating a wide range of levels of accomplishment in non-first language learning. We feel quite confident that the overwhelming majority of South Asians are to at least a minimal extent able to manipulate non-native elements in their linguistic performance. Gumperz has also convincingly shown that for many South Asians the incorporation of even a small amount of "foreign" material, a few key lexical items, the use of particular pronominal forms, may be sufficient to gain acceptability as a speaker of the non-native language.[100] The grammatical structure of discourse produced while code switching is going on may range for successions of discrete complete sentences or sets of sentences, where each is in a unique code, to single sentences with linguistic features of more than one "autonomous code" thoroughly mixed together. The linguistic studies which have been carried out on these phenomena in South Asia are few, but in them and in similar work carried out by linguists in non-South Asian societies, we know that the blending of elements from autonomous codes is not a haphazard process, and can be influenced by complex rules of social and linguistic behavior, and that the ability to manipulate heterogeneous linguistic elements is part and parcel of speakers' competence to linguistically function within an ethnographically complex area. The Stammbaum model of language classification has no meaningful ways of handling such code switching as it is bound by the "uniformity" convention. By this we mean that comparative theory presupposes that at any single time a language can be assumed to have a coherent uniform structure, with units or patterns of units either within the system or outside of it. Multilingual sentences must, therefore, be considered deviant from the point of view of any of the component languages.

5. *Pidginization and Creolization.* A related phenomenon to those just discussed concerns situations of pidginization and creolization. It is well known that numerous "languages" are spoken in South Asia by

individuals who are not native speakers of them. Hindi or Hindustani, for instance, is widely used through north India by speakers who, by any stretch of the imagination are not Hindi native speakers. Prominent examples here are the Baazaar hindustanis of Calcutta and Bombay, as well as the lingua-franca Hindis used in Assam, Jammu-Kashmir, etc. to serve as a link between ethnographically diverse groups. Such languages often contain severe structural simplifications from the grammar of the standard language and are often heavily influenced by the grammatical structure of the native languages of their speakers. When such language varieties are spoken non-natively alongside of actively retained mother tongues they are technically referred to as pidgins. The grammatical structures of pidgins are often highly variable, arising in essence from an attempt to reduce diverse linguistic codes to their lowest common denominator. The specific forms which these compromise codes take are often dependent on the specific structures of the juxtaposed languages and language aptitudes of their users. There is great evidence that such compromise codes are widespread in South Asia, and that much of what passes for Baazaar hindustani, colloquial Nepali, etc., may be the result of pidginization. If it happens that a "pidginized" language variety is learned natively by a second or subsequent generation, then that language variety is generally referred to as a creole. Again there is evidence of widespread creolization in South Asia.[101] Evidence also exists that creolization has played a role in the historical development of at least one modern South Asian language. Southworth has written a number of articles in which he tries, and we think rather successfully, to demonstrate a creole origin for modern Marathi.[102] There are several historical stages in the development of Indo-Aryan which seem to point to creolization, most particularly Buddhist Hybrid Sanskrit and Jaina Ardhamagadhi. South Asian languages exist outside of the subcontinent proper in obviously creolized forms, most notably Carribean Hindi creoles and Fijian Hindi. Moreover, there are numerous language varieties in South Asia, which although not popularly referred to as creoles, seem to fit most of the criteria for them, i.e., Indian English (among those for whom it is native), and Nagamese (possibility of overlap with Pidgins.) No neogrammarian position, it seems to us, is able to account for the origins and many of the structural properties of pidgins. Clearly the origins in such forms lie in the systematic modification of grammatical systems on the basis of interference from other systems. This is a far cry from the internally motivated regular sound laws of comparative historical linguistics.

6. *Irregularities within purely historical models of language families.* It has been known for some time that even from an historical point of view, the Stammbaum model is unable to explain many known points in the evolution of language families. The Stammbaum model necessarily imposes a temporal ordering of language and dialects, and necessitates that earlier reconstructed forms be historically prior to their daughter offshoots. Yet from internal reconstruction alone we know that at virtually all points in the history of certain language families, i.e. Indo-Aryan, there must have been simultaneous use of language at multiple stages in the development of the family. Thus we know, for instance, that at even the earliest periods of our written records of Sanskrit, various middle Indian dialects must have been in common use. This is attested by the presence of clearly unmistakable borrowings from Middle Indic into Old Indic. The Stammbaum model imposes a fossilization of language use which is totally unsupportable from even the textual evidence. At best, then, the Stammbaum conception of the history of Indo-Aryan is a reconstructed record of written documents, and generalizations based on it concerning the history of spoken vernaculars are likely to be misleading. Moreover, the Stammbaum model of language history implies the notion that language diversity is increased with time. One has only to reconstruct far enough back, and one will reach a point at which total uniformity is attained. This is nonsense. There is evidence, for instance, that there were always numerous dialects of Indo-Aryan, at even the earliest periods, and this is attested by the presence in Pali, an early standardized Middle Indic language, of forms which do not correspond to any Old Indo-Aryan (i.e. Vedic) form known to us.[103] There must have been, therefore, other Old Indo-Aryan dialects of which we now have no records, which served as the basis of vernacular dialects which were later standardized into Pali. Similar cases can be shown in the emergence of modern Indo-Aryan vernaculars from various Prakrits and Apabhramsas.

It seems to us desirable to produce methods of analysis which are capable of expressing worthwhile generalizations concerning the phenomena just described. The framework within which Grierson and others operated during the nineteenth and early twentieth centuries, and which has gone largely unchallenged to this day, has great usefulness in providing coherent explanations for many linguistic phenomena. The total framework, however, is a static one, and operates most efficiently when describing relatively homogeneous systems with little social, geographical and stylistic variation. It breaks down completely in coming to grips in situations involving the interaction of "discrete" codes, and those in which there is a great deal of stylistic and social variation. In order to begin to examine the latter kind of phenomena more meaningfully we need dynamic models both of language structure and language change. It is not our place here to construct such models, but we would like to make some suggestions as to what such models must be able to accomplish. 1. They must be able to describe the full range of linguistic variation used by an individual in a full range of social contexts. Such linguistic competence may very well encompass what in traditional terms would be thought of as the mixture of two or more codes. In other words, such a model must be able to describe in Uriel Weinreich's terms, "the linguistic repertoire of individuals." In cases where the speech of individuals involves structured variability, it must be able to enumerate the constraints on this variability. Past experience in sociolinguistic theory has indicated that some of these constraints are likely to be purely linguistic and some social. 2. The model must be able to generalize beyond the linguistic competence of the individual to the cumulative linguistic competence of social groups. Where groups control heterogeneous linguistic systems, and functional importance to the use of elements of these systems, the model must be able to account for the conditioning factors determining the selection of particular components of complex systems. In other words, again using Weinreich's terms, the model must be able to give form to the notion of "diasystem", that is, a formal linguistic system describing both the structural properties shared by the members of a group and the structural linguistic properties used by particular subsets of the group.

NOTES: CHAPTER 2

1. For an excellent overview of the literature on just the work done in this area in the United States over a brief period of time, see Schofield Andrews Jr. and Joshua Whatmough, "Comparative and Historical Linguistics in America: 1930-1960," in Christine Mohrmann, Alf Sommerfelt and Joshua Whatmough (eds.), *Trends in European and American Linguistics: 1930-1960,* (Utrecht and Antwerp: Spectrum Publishers, 1966), pp. 58-81.

2. For a detailed examination of the early development of dialectological studies see Sever Pop, *La Dialectologie: Aperçu Historique et Methodes d'Enquêtes Linguistiques.* 2 volumes. (Louvain: Chez l'Auteur, 1950). See also Jaberg, 1936.

3. ". . . a set of isoglosses running close together in much the same direction--a so-called *bundle* of isoglosses--evidences a larger historical process and offers a more suitable basis of classification than does a single isogloss that represents, perhaps, some unimportant feature." (Bloomfield, 1933:324).

4. Such a task would obviously be impossible. Dialect geography has commonly limited itself to phonological, lexical, and morphological variation.

5. See particularly Kurath, 1972 and Lehmann, 1962, pp. 115-46.

6. For an excellent example of this position see H. A. Gleason, Jr., *An Introduction to Descriptive Linguistics*, revised edition (New York: Holt, Rinehart, and Winston, 1961), pp. 401-3.

7. Cf. Gleason, *ibid.*, section 24.14, p. 400.

8. *Ibid.*

9. It is, of course, impossible to demonstrate total non-relatedness of

any two speech varieties. They will necessarily show some linguistic features--be they partial phonological inventory, a class of grammatical categories, or lexical items fortuitously similar in phonological shape--in common.

10. Thus the most prominent work in dialect geography in the west has been carried out with regard to English, German, Italian, French, Romanian, etc., all languages whose essential existence *qua* independent languages is not seriously open to question.

11. This point has been effectively made in Stankiewicz, 1957 and Weinreich, 1954.

12. For a detailed history of the development of Indo-European historical linguistics, see Pedersen, 1959.

13. The development of this model is generally credited to August Schleicher, under the inspiration of F. Ritschl. (F. Robins, 1967:178-80; see also J. P. Mayer, "More on the History of the Comparative Method: The Tradition of Darwinism in August Schleicher's work," *Anthropological Linguistics* 8(1966):1-12.

14. Cf. Bloomfield, 1933:297-320.

15. *Ibid.*

16. See E. Pulgram, "Neogrammarians and Sound Laws," *Orbis* 4(1955):61-5.

17. See Bloomfield, 1933:281-96.

18. Few if any linguists, to the best of our knowledge, ever explicitly stated that it was possible to ascertain the structure of a fully homogeneous proto-language through this method. Yet by constructing inverted tree diagrams in which the uppermost node represented the earliest ancestral member of a family of languages, they lent a greater degree of concreteness to their reconstructions than might otherwise have been warranted.

19. Cf. Pedersen, 1959:311-39.

20. The resemblance may, at times, be tenuous. Often one finds modern standardized vernaculars assigned such a position in classification charts. There may be a considerable gulf between the structures of these standardized vernaculars and the structures of their spoken conversational varieties.

21. "The fact that [dialectological] differences themselves follow a system,--that the difference, say, of English and German [aw] and Dutch [øy] appears in a whole series of forms--confirms our surmise that these forms are historically connected. The divergence, we suppose, is due to characteristic changes undergone by some or all of the related languages." (Bloomfield, 1933:300).

22. This chart, adapted from Schleicher, still imposes an overly great deal of binary divisions into the history of Indo-European languages, and accepts the postulation of a unified Proto-Indo-European language.

23. Primarily Johannes Schmidt. See Bloomfield, 1933, section 18.12, pp. 317-8.

24. *Ibid.*

25. Cf. Gleason, *op. cit.*, pp. 401-3.

26. That is to say, the results of the operation of a single sound law may join the class of sounds which undergo subsequent laws. The cumulative effect of such changes over time can be considerable.

27. Thus, for example, the modern Romance languages emerged through the according of political and social recognition to those speech varieties arising through the force of a series of phonological laws applying to Old Italic speech varieties.

28. This latter is essentially Schleicher's position.

29. Cf. Robins, 1967:179.

30. Such a position allows linguistic differentiation to take place while linguistic contact is maintained by individuals or social groups. This is a far cry from the claim that linguistic innovation occurs as a function of a lack of contact between groups (i.e. the position that linguistic homogeneity is increased by linguistic contact).

31. Cf. Robins, 1967:179.

32. See particularly Bailey, 1972, 1973a and 1974b, as well as the papers in *NWAVE*.

33. We return to the North Indian situation in Chapters 3 and 5.

34. For information on the development of structuralism in linguistics see Francis P. Dinneen, S. J., *An Introduction to General Linguistics* (New York: Holt, Rinehart and Winston, 1967) pp. 192-98. See also C. F. Voegelin and F. M. Voegelin, "On the History of Structuralizing in 20th Century America," *Anthropological Linguistics* 5:12-35.

35. For further analysis of Saussure's views on language, see Charles Bally, *Ferdinand de Saussure et l'état actuel des études linguistiques* (Geneva, 1913); Dinneen, *op. cit.*, pp. 192-212; R. Godel, "F. de Saussure's Theory of Language," in *CTL5*, pp. 479-93, and Rulon S. Wells, "De Saussure's System of Linguistics," *Word* 3(1947):1-31.

36. This doctrine is well illustrated in Part Two of the de Saussure *Course*. [de Saussure, 1959:101-39]

37. Cf. Dinneen, *op. cit.*, pp. 195-6.

38. Cf. Pulgram, *op. cit.*

39. Cf. Philip W. Davis, *Modern Theories in Language* (Englewood Cliffs, New Jersey: Prentice Hall, 1973), pp. 14-38.

40. See particularly Stankiewicz, 1957 and Weinreich, 1954.

41. Robins, 1967:190-1.

42. Cf. Weinreich, 1954.

43. These terms when used in these senses were first popularized by de Saussure. Cf. Dinneen, *op. cit.*, 205-6; de Saussure, 1959:122-7.

44. This notion too was most widely espoused by de Saussure. Cf. de Saussure, 1959:79-100.

45. Such a distinction would be untenable in the case of many South Asian languages, where such clearly distinct languages such as Hindi, Bengali, and Gujarati will show common syntactic processes, but differ in many phonological rules.

46. Particularly Greenberg, 1956.

47. See particularly "The Phonemic and Grammatical Aspects in their Interrelations," in *Selected Writings, II* (The Hague: Mouton, 103-14) and *Zur Struktur des russischen Verbums* in *Selected Writings, II*, pp. 3-15.

48. Weinreich, 1954.

49. See Jacobson, Fant, and Halle, 1951, for a discussion of a set of "distinctive features" usable for describing all human languages.

50. For a discussion of neo-Bloomfieldian phonology, see Davis, *op. cit.*, 128-72 and Bernard Bloch, "A Set of Postulates for Phonemic Analysis," *Language* 24(1948), 3-46.

51. We make no attempt to sort out here the numerous schools of transformational-generative grammar which have developed in the past year, as many of these schools have not directly concerned themselves with matters of dialectal variation.

52. This "competence/performance" dichotomy was clearly articulated in Noam Chomsky's *Syntactic Structures*, The Hague: Mouton, 1957, as well as in many of his later writings.

53. See particularly Uriel Weinreich, *Explorations in Semantic Theory* (The Hague: Mouton, 1973). Dissatisfaction with the conception of a semantic component found in Chomskian generative grammar is also one of the primary motivations in the recent development of "generative semantics," a school of linguistics one of whose major goals has been the development of abstract levels of linguistic representation which more adequately represents the semantic properties of real-world linguistic utterances.

54. ". . .the grammatical description of a given dialect may be converted into an adequate description of a related dialect by the addition, deletion, or reordering of a relatively small number of rules. Indeed, it is tempting to propose that the degree of difference between dialects is nothing more than a function of the number and type of such changes." (Saporta, 1965:219)

55. This paradox is well described in Weinreich, Labov, and Herzog, 1968.

56. Virtually all introductions to linguistics begin with a statement to the effect that language is fundamentally a social entity, but procede to state that it is only by systematically disregarding this fact that linguistic description is at all able to be carried on. Gleason, in his *Introduction to Descriptive Linguistics*, for instance, states that

> language has so many interrelationships with various aspects of human life that it can be studied from numerous points of view. All are valid and useful, as well as interesting in themselves. Linguistics is the science which attempts to understand language from the point of view of its internal structure. (Gleason, *op. cit.*, p. 2)

57. Most particularly from the many writings of Hugo Schuchardt, who dealt with many of the same concerns of modern social dialectologists.

58. For a history of dialect geography in the west, see Bottiglioni, 1954.

59. By which is to be understood that the data were not in accord with the prestigious prescriptive norms of the languages.

60. Such an ability constitutes neither linguistic "competence" or "performance" in the traditional sense of those terms, but rather what we can term "sociolinguistic competence."

61. For a further discussion of Hindi relative clause structure, see Susan K. Donaldson, "Movement in Restrictive Relative Clauses in Hindi," In Yamuna Kachru (ed.), *Papers on Hindi Syntax* [= *Studies in the Linguistic Sciences*, vol. 1, no. 2 (Fall, 1971)], pp. 1-74.

62. *Ibid.*

63. For a discussion of the notion "inherent variability", see William Labov, "Contraction, Deletion, and Inherent Variability of the English Copula," in *Language in the Inner City* (Philadelphia: University of Pennsylvania Press, 1972), pp. 65-129.

64. *Ibid*, p. 73.

65. Wolfram and Fasold, 1974, p. 162.

66. *Ibid*, pp. 124-76.

67. *Ibid*, pp. 129-34.

68. Henrietta J. Cedergren and David Sankoff, "Variable Rules: Performance as a Statistical Reflection of Competence." *Language* 50 (1974), 333-55. See also Cedergren, 1973.

69. William Labov, "Stages in the Acquisition of Standard Language," in *Social Dialects and Language Learning*, ed. by Roger W. Shuy. Champaign, Ill.: National Council of Teachers of English, 1964, 102.

70. Quoted in Robins, 1967:134.

71. S. H. Kellogg, *A Grammar of the Hindi Language*, 3rd ed., (London, 1875). Reprinted in London: Rutledge and Kegan Paul, 1955.

72. John Beames, *A Comparative Grammar of the Modern Aryan Languages: To Wit, Hindi, Panjabi, Gujarati, Marathi, Oriya and Bengali* (London: Trubner and Co., 1872-79). Reprinted in Delhi, India: Munshiram Manoharlal, 1966.

73. Turner, 1966.

74. *LSI*, vol. 1.

75. Sir George A. Grierson, *On the Modern Indo-Aryan Vernaculars* (Bombay: 1933).

76. Anquetil du Perron, *Zend-Avesta, ouvrage de Zoroastre*. 1771.

77. A. V. W. Jackson. *An Avesta Grammar in Comparison with Sanskrit*. 1892. Reprinted 1968.

78. H. Reichelt. *Awestisches Elementarbuch*. (Heidelberg: Winter, 1909). Reprinted 1968.

79. Christian Bartholomae. *Altiranisches Wörterbuch* (Strassburg, 1904; reprinted Berlin, 1961).

80. Geiger and Kuhn, 1901.

81. A. D. Campbell. *A Grammar of the Teloogoo Language* (Madras, 1816). Third edition Madras, 1849. Note to the introduction by Francis W. Ellis, pp. 1-31.

82. R. Stevenson, "A Comparative vocabulary of non-Sanskrit vocables of the Vernacular languages of India,' *Journal of the Bombay Branch of the Royal Asiatic Society* 5 (1852).

83. Caldwell, 1856.

84. Burrow and Emeneau, 1961.

85. Krishnamurti, 1969a.

86. K. Zvelebil. *Comparative Dravidian Phonology.* (The Hague: Mouton, 1970).

87. Pinnow, 1959.

88. Especially Zide, 1969 and Zide and Stampe, 1968.

89. Pinnow, 1963.

90. Kuiper, 1948.

91. Miller, 1969.

92. Shafer, 1949, 1953, 1955 and 1960.

93. Uray Géza, "Kelet-Tibet nyelvjárásainak osztályozása" [Classifications of the dialects of eastern Tibet], *Bölcsészetkari értekezések* 17, Dissertationes falcultatis philosophical, Universitatis scientiarum a fundatore Petro Pázmány nominatae (Budapest, 1949).

94. A. Róna-Tas, *Tibeto-Mongolia, The Tibetan Loanwords of Monguor and the development of the archaic Tibetan dialects* (Budapest, 1966).

95. Particularly Nishida Tatsuo, "Birumago to Roro shogo, sono seichō taikei no hikaku kenkyū" [Burmese and the Lolo languages, their tone systems and comparative studies], *Tōnan Ajia Kenkyū* 4 (1964), 13-28.

96. *LSI*, vol. 1, p. 16.

97. See Chapter 4 in this volume.

98. Pandit, 1963.

99. Gumperz, 1958, 1964a, b.

100. Gumperz, 1964a, b.

101. For a partial bibliography of pidginization and creolization in South Asia, see John Reinecke, *et al.*, *A Bibliography of Pidgin and Creole Studies* (Honolulu: The University Press of Hawaii), 632-5.

102. Southworth, 1971.

Chapter 3

Traditional Taxonomies of South Asian Languages

3.0. Introduction

In the previous chapter we attempted to demonstrate that there is no universally accepted set of criteria by which one can group related speech varieties into languages and dialects. The terms "language" and "dialect" themselves have no uniform meanings, and what is understood by them is a function of the theoretical and methodological approaches and unstated assumptions of the individuals using them. There is little agreement as to how these terms may be used to characterize particular speech forms found in different parts of the world. Those areas of the world where there has been the least controversy as to the identification of what constitutes "autonomous languages" have largely been those where the criteria used by different groups of schools of linguists have fortuitously coincided. Such circumstances have arisen generally when factors of nationality, literary traditions, and language consciousness conspire to produce the common acceptance of terms to refer to specific linguistic entities. In other areas of the world, with South Asia a prominent example, there has not been such a clear overlapping of criteria, and it is substantially more difficult to gain a consensus as to the use of labels in referring to different orders of speech varieties.

The thrust of these claims is to assert the ultimate impossibility of linguistically defining what is meant by the terms "language" and "dialect". The most that can be expected of linguistics is that its practitioners make every attempt to spell out in their works precisely what aspects of linguistic usage they aspire to describe. Situations in which there is widespread agreement as to the existence of certain languages, "i.e. French, Russian, Gujarati," arise through the interaction of social, political, historical and psychological factors and are not reflections of any inherent properties of the linguistic "entities" referred to by these terms.

With this said, we would like to discuss in this chapter the traditional taxonomy of languages and dialects in South Asia. As the major emphasis in this report is on socially conditioned language variation in the region, the inclusion of this chapter requires some comment. In probably few areas of the world is the characterization of the distribution of language varieties as confusing as it is in South Asia. While numerous terms exist which are commonly said to refer to South Asian languages (e.g. Marathi, Hindi, Sanskrit, etc.) there is little agreement as to exactly what these terms refer to. It is virtually impossible to establish a clear vertical hierarchy of superordinate "languages" and subordinate "dialects" anywhere within the region. The speech forms within the region differ markedly from one another in their degree of standardization, length and substance of literary traditions, and the degree of psychological allegiance which they command from their speakers. The linguistic competence of many individuals encompasses elements from what is, in fact, more than one of these speech varieties, and the conditions under which they switch back and forth among these diverse components is not clearly understood. Many South Asian speech varieties are referred to by several names, and the names of many South Asian speech forms--Hindi, Hindustani, Urdu notable among them--are used by writers to refer to a wide variety of linguistic entities and for a myriad of purposes. It is clearly impossible to sort out here all of the ambiguity inherent in referring to South Asian speech forms, and going one step further, we believe that it may even be theoretically impossible to do so.

Nevertheless, we feel that it is imperative to have some frame of reference within which one can talk about language use in South Asia. Attempts have been made for at least two hundred years to arrange South Asian speech forms into coherent typological systems. Almost invariably the criteria which have been used in doing this are those that were employed in 19th century comparative and historical linguistic studies. While there is certainly nothing inherently wrong in such frameworks, they have produced a massive amount of highly systematic data on the Indo-European languages, and have necessarily made it difficult to formulate generalizations about the types of sociolinguistic variation which we are interested in here. In discussing the taxonomies of South Asian languages from a traditional point of view, we do not therefore wish to create the impression that the pattern of linguistic distribution in South Asia which they provide is an optimally correct one, but rather that this pattern provides a ready starting point for the examination of language use in the region from other points of view.

In Section 3.1 we discuss the Indo-Aryan languages of South Asia, describing a few of the most common statements of their mutual relations, and enumerating what have commonly come to be considered the major dialects of a number of them. In Section 3.2 we turn our attention to the Dravidian languages and discuss each of a number of attempts to determine the internal relations among that group of languages. The construction of an overall hierarchy of the Dravidian languages has, in general posed fewer problems than that of the Indo-Aryan languages, and, as we shall see, much of the recent work on Dravidian typology has focussed on integrating the ever-increasing corpus of data on the non-literary Dravidian languages into the overall skeleton of relations with that language family. In Section 3.3 we offer a cursory discussion of the relations among the so-called Munda languages, and Section 3.4 examines attempts to construct a taxonomy of Tibeto-Burman languages of South Asia.

It is obviously impossible in a survey of this sort to present a comprehensive study of the history of taxnomic studies of all four of these language families, leaving aside a full description of the specific linguistic data which have been utilized in these studies. Our survey is, therefore, a brief one, and no attempt has been made to thoroughly summarize the literature on South Asian language taxonomy. The sections differ in the quantity and comprehensiveness of the discussions. The material on the Indo-Aryan languages is largely derivative of a number of standard sources, and no attempt has been made to exhaustively list all of the many "dialects" of the languages of this family. Only overall typological schema have been presented. The material on the Dravidian languages is somewhat more extensive than that concerning the Indo-Aryan languages, and we have attempted at least a rapid survey of the most important typological systems for the language family. The sections on Tibeto-Burman and Munda languages are quite perfunctory, and taken directly from a small number of published sources, with little attempt made at integrating basic typological data or reconciling differences.

3.1. Classification of the Indo-Aryan languages

3.1.0. Introduction

The existence of an autonomous group of Indo-Aryan languages, whose approximate relationship to Sanskrit was assumed to roughly parallel that of the modern Romance languages in relation to Latin, has been known for hundreds of years.[1] There has been little doubt concerning the existence of a familial relationship holding between Sanskrit and a large number of the vernacular languages of North India, Pakistan, Bangladesh, Nepal, as well as Sinhala, spoken in Sri Lanka. While the existence of this relationship has appeared obvious for some time, the specification of its precise internal structure has remained exceedingly difficult. There are a number of reasons for this set of circumstances. First of all, unlike the situation in modern Europe, it is virtually impossible to precisely specify, even on a purely synchronic level, which are the standard vernacular languages of the area. Thus it is well near impossible to answer such questions as: should Konkani be considered a dialect of Marathi or an autonomous language having a position parallel to it?; is Rajasthani [leaving aside the problem of defining that entity] a dialect of Hindi or an autonomous language parallel to it?; is Maithili an autonomous language, a dialect of Hindi, or one of Bengali?; and are Hindi and Urdu one language or two? Within a purely comparative typological system the ability to answer these and similar questions requires the ability to reconstruct unified proto-systems from which sets of daughter languages or dialects may be derived by regular rules. This is rendered extraordinarily difficult by (1) the absence of written records for many of the earlier stages of the language families; (2) an amazing proliferation of dialectal variants in the spoken vernacular, these vernacular dialects often in close proximity to one another and mutually influencing each other, the resultant pattern of variation not being easily amenable to historical and comparative analysis; and (3) the obliqueness of the relationship holding between the standard earlier languages for which we do have records and the spoken vernaculars which are presumably contemporary with the standardized languages and typologically subordinate to them. The standardized Prakrits and Apabhramsas were often highly stylized languages, considerably removed from what we know of the structures of these vernacular languages, and drawing much of their technical vocabulary from sources external to the vernacular context. These languages are often considerably conservative at even the times at which they are first used, and reconstruction attempts to relate later historical stages of a language family, which are likely to have their origins not in standardized speech varieties, but in unrecorded spoken ones, are apt to be misleading. In addition to these factors, it is important to note (4) that in many cases the emergence of standard modern vernacular Indo-Aryan languages dates only to the last few centuries, and has involved conscious processes of relexification and borrowing, these obscuring the historical affinities among the language families. For political, social, and other non-linguistic reasons, many earlier Indo-Aryan vernaculars which potentially might have become modern standard languages did not do so.[2] The reconstruction of a family history based only on those which did is impossible. A last feature rendering the classification of the Indo-Aryan languages difficult is (5) the extreme paucity of data on many of the non-standard language varieties which play a pivotal role in the construction of an adequate Stammbaum for the family. The first and only systematic attempt to gain a large body of data upon which to base a classificatory analysis of the Indo-Aryan languages was of course, done by Grierson prior to the publication of the *LSI*. Yet as considerable as the amount of data gathered by Grierson is, it is nevertheless a fraction of what is required. Moreover, Grierson's data has as yet been followed by few subsequent efforts and it is not unfair to say that the classificatory attempts which have taken place since his time have, in one way or another, involved rehashing of his data.

In this section we examine the overall schema for the Indo-Aryan languages given in a small number of sources (Beames, 1960 [original, 1866-7], Grierson in *LSI*, vol. 1, Chatterji, 1926). Material from some other sources is cited in expanding what those three sources provide about certain specific languages and dialects.

3.1.1. Beames on Indo-Aryan

In reading Beame's characterization of the Indo-Aryan languages, one is struck by how little effort is made to enumerate the specific speech forms to be included within the family. The existence of an "Aryan" family of languages is assumed by Grierson, and the problems which pose greatest interest for him are those involving the formulation of rules to explain similarities in vocabulary and inflexion among the literary members of the family. The unifying factor among the members of the family is the fact that:

> Spoken Sanskrit is the fountain from which the languages of Aryan India originally spring; the principal portion of their vocabulary and the whole of their inflectional system being derived from this source. Whatever may be the opinions held as to the subsequent influences which they underwent, no doubt can be cast on this fundamental proposition. Sanskrit is to Hindi and its brethren, what Latin is to Italian and Spanish. (Beames 1960:2)

Beames holds the analogousness of the Aryan languages to the Romance languages quite seriously. He maintains that the modern Aryan languages all maintain the distinction between lexical items which have been borrowed from Sanskrit fairly recently, the so-called *tatsama* items, and those which have been derived from Sanskrit through the cumulative effect of historical sound laws, the so-called *tadbhava* items. All of the modern Aryan languages, according to Beames, agree in maintaining this distinction, but differ in the degree to which either *tatsama* items or non-Aryan synonyms supplant the *tadbhava* lexicon. Beames goes to considerable length to document the incorporation of non-Italic items to Vulgar Latin, supplanting etymologically pure classical Latin ones.

Beames considers the raw material for the comparative analysis of the Indo-Aryan languages to be essentially *tadbhava* lexical items and basic morphology. He places particular importance on the role of *tadbhavas* in determining linguistic relatedness:

> It is to the Tadbhavas that we must turn if we would become acquainted with the secrets of the phonetic machinery of the Aryan Indians. Of these there are two sorts, so distinguished from one another that it is impossible to mistake them. The one class consists of those words which were in use in Prakrit, and in which the Prakrit processes have been carried one step further. The other contains words which apparently have not come through Prakrit, as they exhibit a more perfect form, and a nearer approach to the Sanskrit than the Prakrit form does. (Beames 1960:13-14)

It is vocabulary, without a doubt, that is of central interest to Beames in describing the relations among Indo-Aryan speech forms. The selection of seven languages to serve as the foundation for the comparative analysis of the family, to wit, Hindi, Punjabi, Sindhi, Gujarati, Marathi, Oriya and Bengali seems somewhat arbitrary, and no doubt influenced by the existence of standardized varieties and literary traditions for these speech forms. Beames rather strangely considers these to be a hierarchy of sorts among these seven and refers to Hindi, Marathi and Bengali as "the three principal [languages]" (p. 31), with Hindi enjoying a superior position among these, being called "the legitimate heir of the Sanskrit, and fill[ing] that place in the modern Indian system which Sanskrit filled in the old." (p. 31) In structuring these seven speech forms into a typological system, Beames uses the degree to which each possesses Sanskritic *tatsama* vocabulary, *tadbhava* items, and Persian Arabic loan words. By this criteria, the seven form the following system:

Panjābī		Hindī		Baṅgālī
Sindhī	Gujarāṭī		Marāṭhī	Oṛiyā

Figure 27. The Indo-Aryan languages (from Beames, 1960:40).

In this system the left side of the page represents the "Perso-Arabic" pole, the right side the Sanskritic pole, and the proximity on the printed page of a language to each pole represents the extent to which the vocabulary of that language contains vocabulary items of the three enumerated types.

In spite of Beames' declared pro-Hindi orientation, and in spite of the limited value of his taxonomic attempts, he does provide useful insights on the dissemination of the Indo-Aryan languages of North India. Beames notes that the Indo-Aryan area is generally agreed to exhibit a chain of mutually-intelligible speech forms beginning with Sindhi in the West and extending to Bengali in the East, Marathi in the South, and Kashmiri in the North. No abrupt dialect boundaries are to be discerned between any speech form and the immediately contiguous one, so that a situation very much like the Romance language area can be said to exist. Within various areas at various times in recent modern history, literary forms of various speech forms have arisen, some of which have continued to this day to be used as official languages (media of instruction, books, newspapers) in various areas. Beames recognizes seven of these modern Aryan languages (Hindi, Panjabi, Sindhi, Gujarati, Marathi, Oriya, and Bengali)[3] but repeatedly makes statements to the effect that ". . .the various cognate languages of Aryan India melt into one another so imperceptibly that it is a question of much difficulty how to define their limits." (Beames 1960:99)

Some of the Indo-Aryan dialects have names, sometimes because they had literary forms at one point but have since ceased to be used for literature, e.g., the Hindi 'dialects' Magadhi, Maithili, Bhojpuri (which Beames notes retain "many fine old Aryan forms which have dropped out of classical Hindi,") (Beames 1960:97) Marwari, and others.

This last, Marwari, spoken in Rajputana, Beames finds hard to distinguish from Marathi, Panjabi, and Sindhi. The boundaries of the dialects of Panjabi, Beames notes, are themselves difficult to determine, and the enumeration of the various Panjabi dialects is also impossible due to the "absence of any written standard." In fact, Beames believes Panjabi to have more dialects per unit of area than any other level territory in India. For Beames, presence of a written standard helps in the definition of language.

Beames attempts in his work to enumerate what he considers to be the major dialects of each of the major languages which he treats, and we briefly summarize here his treatment of these dialects.

For Sindhi, following Trumpp, Beames enumerates three dialects: Sirai, Vicholai, and Lari, but notes that more variations exist. Gujarati tends to merge with Sindhi--the Kacchi dialects can be said to be either Gujarati or Sindhi, or half one and half the other. In the North and Northeast Gujarati merges with Vraj Marwāri, or Hindi. In the South, however, the boundary with Marathi seems to be more abrupt, in that Gujarati and Marathi speakers claim to not understand one another. Beames thus feels Gujarati to have a close relationship with languages to the north of it, and makes some historical claims for such a northern origin of Gujarati.

Marathi, according to Beames, has two divisions--Konkan and Dakhani. The former is spoken on the coast, the latter, inland. The Poona dialect, called Deshi, is considered standard. In the southern Dakhan, the language is mixed with Canarese, Beames notes, while Christians mix in much Portuguese. Beames disclaims first-hand knowledge of Marathi, however, never having been stationed in that area.

Oriya, Beames claims, is the most homogeneous of the modern Aryan languages, except for in the north where it tends to get mixed with

Bengali; the 'purest' Oriya is to be found spoken in the hills.

Bengali, by contrast, displays a bewildering crowd of dialectal forms, with little uniformity aside from the Calcutta standard. Apparently the use of Calcutta standard (Dimock: Literary and Colloquial Bengali) by educated Bengalis as a lingua franca in their area had not spread very far in Beames' time. Beames distinguished an Eastern, Northern, and Southern dialect of Bengali, and specifies phonological features of this differentiation.

Beames considers the dialect situation of Hindi to be of particular importance, as he accords that language a position of preeminence in Indo-Aryan, and goes to some length to spell out what he means when he refers to it:

> Hindi is that language which is spoken in the valley of the Ganges and its tributaries, from the watershed of the Jamnâ, the largest and most important of them, as far down as Rajmahal, the point where the Ganges takes a sudden turn to the south, and breaks out into the plains of Bengal. This area is the centre and principal portion of Aryan India. It includes the Antarbed or Doab between the Ganges and the Jamnâ, the "inner hearth" of the nation. It is therefore the legitimate heir of the Sanskrit; and fills that place in the modern Indian system which Sanskrit filled in the old. Under the general head of Hindi are included many dialects, some of which differ widely from one another, *though not so much so as to give them the right to be considered separate languages* [emphasis added: MCS and HFS]. Throughout the whole of this vast region, though the dialects differ considerably, one common universal form of speech is recognized, and all educated persons use it. This common dialect had its origin apparently in the country round Delhi, the ancient capital, and the form of Hindi spoken in that neighborhood was adopted by degrees as the basis of a new phase of the language, in which, though the inflections of nouns and verbs remained purely and absolutely Hindi, and a vast number of the commonest vocables were retained, a large quantity of Persian and Arabic and even Turkish words found a place, just as Latin and Greek words do in English. Such words, however, in no way altered or influenced the language itself, which, when its inflectional or phonetic elements are considered, remains still a pure Aryan dialect, just as pure in the pages of Wali or Saudâ, as it is in those of Tulsi Dâs or Bihâri Lâl. (Beames 1960:31-2)

In addition to considering Hindi to be essentially a single "language" encompassing a number of dialects, Beames accords this language linguistic superiority over the other standard Indo-Aryan languages:

> All of the other languages of the group were originally dialects

> of Hindi, in the sense that Hindi represents the oldest and most widely diffused form of Aryan speech in India. Gujarati acknowledges itself to be a dialect of the Sauraseni Prakrit, the parent of Hindi. Panjabi, even at the present day, is little more than an old Hindi dialect. Bengali, three centuries ago, when it first began to be written, very closely resembled the Hindi still spoken in Eastern Bihar. Oriya is in many respects more like Hindi than Bengali. . . . (Beames 1960:33)

It is clear then that Beames is as much interested in justifying the existence of a vertical hierarchy of modern Aryan languages, with a unified Hindi occupying the top-most node, as he is providing a purely descriptive taxonomy of the spoken vernaculars of North India. He operates with a number of unstated but nevertheless evident criteria by which he is able to distinguish between "languages" and their "dialects":

(1) speech forms which have a modern literature are languages;

(2) named or literary dialects (no longer used as modern languages) are more important, and rank higher on some kind of scale, than unnamed dialects;

(3) geographical dialects are not to be distinguished from social dialects in any systematic way;

(4) pidginized or creolized languages (e.g. Konkani with Portuguese vocabulary) have less status than "pure" dialects which are lexically conservative;

(5) historically more conservative dialects (morphologically, etc.) with "fine old Aryan forms" are the bluebloods of the Indo-Aryan scene, while innovative dialects and languages are to be stigmatized.

3.1.2. Grierson on Indo-Aryan.

Perhaps the greatest contribution to Indo-Aryan dialectology is the *Linguistic Survey of India* (Grierson, 1903-28). Grierson's analysis contrasts strongly with Beames' (where everything merges imperceptibly from one area into another) in that Grierson divides the modern Indo-Aryan languages into two groups: a central or Mid-land (Madhyadēśa) group, surrounded by an outer ring of languages/dialects "beginning in Hazara in the Panjab, and running through the Western Panjab, Sindh, the Maratha country, Central India, Orissa, Bihar, Bengal and Assam." (Grierson, 1903-28:117) The main difference between the inner core and outer ring seems to be based on the treatment of /s/, which remains /s/

only in the inner core, but is changed to /h/, /sh/, or even a palatal fricative in the languages of the outer core. This is a difference Grierson believes to exist since the time of Herodotus. There are also differences in the declension of nouns and the conjugation of verbs--the inner core being analytic, while the outer ring has gone from "synthetic" through "analytic" into a new "synthetic" stage. Grierson constantly talks of historical development from Sanskrit, the outer ring being descended from dialects of Sanskrit which differed from those dialects of Sanskrit from which the inner core descended. A typical example of this is the development of enclitic pronouns in the outer ring, where they have developed into personal terminations, while in the inner core this has not happened.

Grierson subdivided the three subgroups (it turns out that a 'mediate' sub-branch consisting only of eastern Hindi must also be taken into account) into the following languages:

- A. Outer sub-branch
  - I. Northwestern Group
    - 1. Lahadā or Western Pañjābī
    - 2. Sindhī
  - II. Southern Group
    - 3. Marāṭhī
  - III. Eastern Group
    - 4. Oṛiyā
    - 5. Bihārī
    - 6. Bengali
    - 7. Assamese
- B. Mediate Sub-branch
  - IV. Mediate Group
    - 8. Eastern Hindī
- C. Inner Sub-branch
  - V. Central Group
    - 9. Western Hindī
    - 10. Pañjābī
    - 11. Gujarātī
    - 12. Bhīlī
    - 13. Khāndēśī
    - 14. Rājasthānī

VI. Pahāṛī Group

15. Eastern Pahāṛī or Naipālī

16. Central Pahāṛī

17. Western Pahāṛī

Unspecified.

Grierson specifies that "of the above, Marathi and Eastern Hindi are groups of dialects, not of languages. The languages of the Pahaarii Group are those spoken in the lower Himalayas." (Grierson, 1903-28:120)

Aside from the question of phonological and grammatical developments in the outer ring which differ from those in the inner core, Grierson also uses the criterion of mutual intelligibility of adjacent dialects to determine subgroupings. For instance, although Marathi is a member of the outer ring, and merges with Oriya (also outer ring) in the east, it does not shade off into Western Hindi, Gujarati (formerly an outer ring language but now superimposed by an inner core language), or Sindhi. However, it merges gradually with Eastern Hindi, and since Eastern Hindi merges gradually with Western Hindi, this makes Eastern Hindi anomalous in its status as neither inner or outer.

Clearly the importance of the *LSI* with regard to our understanding of Indo-Aryan is in its attempt to elicit a body of primary linguistic data in as wide a body of spoken vernaculars as possible, to provide texts in those vernaculars, and to use these bodies of data as the basis of typological statements about the language family. Once the data were collected, the historical evolution of the language family could be outlined in purely historical terms, tracing the evolution of the modern Indo-Aryan languages from Sanskrit through the Prakrits and Apabhramsas.[4] Grierson and his associates collected a huge body of data, and in interpolating these data Grierson necessarily made certain assumptions about how groups of dialects are to be grouped under the rubric of single languages. There are seldom explicit statements as to the criteria used in doing this, and it is necessary to examine statements about particular speech forms to determine what Grierson's methodological assumptions were.[5]

Some quotes from Grierson may be illustrative of his theory of dialectology: "Returning to the Bombay Presidency, we must consider one form of Marāṭhī which is a real dialect, and not merely a corrupt form of the standard form of speech. This is KōnkaṆī. . . ." (Grierson, 1903-28:144)

Grierson goes on to say that Konkani is a "real" dialect because it diverged early from Marathi, and once had a literature, and in fact is written in a number of different writing systems in different areas. This is contrasted with some areas in Orissa, where some "mixed dialects, half-Oriya and half-Bengali" are spoken. (Grierson, 1903-28:146) "Of these there are almost as many forms as there are speakers, the two languages being mixed at random according to the personal equation of each. A sentence may begin in Oriya and end in Bengali. . . but all this does not constitute any definite dialect. . . ." (Grierson, 1903-28:146) Grierson may have been observing code-switching here. Other examples of Grierson's dialectology, on Marathi dialects:

> No less than 39 names have been recorded in the Survey as those of dialects of Marāthī. Few of these can be called genuine dialects, the majority being merely forms of the standard speech or of one of the real dialects, pronounced in some peculiar way according to locality or to the caste of the speakers. For instance, the Marāṭhī of the Konkan north of Ratnagiri is very nearly the same as the standard, but natives recognize two dialects, one spoken by the Brahmans, and another spoken by Musalmans. These minute differences are all investigated in the pages of the Survey, but here would be manifestly out of place. It will be sufficient to mention here the four main dialects, viz., Dēśī, Konkan Standard, the Marāthī of Berar and the Central Provinces, and Kōṅkaṇī." (Grierson, 1903-28:145)

and also, on the Berar dialect:

> Historically, it [the Berar dialect, called Verhāḍī] should represent the purest Marāṭhī for Berar corresponds to the ancient Vidarbha or Mahārāshṭra; but in after centuries the political centre of gravity moved farther west, and with it the linguistic standard. (Grierson, 1903-28:144)

Also on Hal^{a}bī, also called Bastarī:

> Hal^{a}bī, also called Bastarī. . . , was for long nobody's child in the linguistic classification of India. Our Survey shows that it is a corrupt mixture of several languages, both Aryan and Dravidian, forming a transition tongue between Marathi and Oriya, but generally with a Marathi backbone. . . . Returning to the Bombay Presidency, we must consider the one form of Marathi which is a *real* dialect, and not merely a *corrupt* form of the standard form of speech. This is Konkani. . . . As a dialect of Marathi, it branched off from the common parent Prakrit at a relatively early period, so that there are many divergencies from the standard of Poona. (Grierson, 1903-28:144)

Grierson's treatment of Bihari shows the same bias toward speech forms which may have once had a written form to be considered as "dialects," whereas those which show admixture or convergence with other speech forms are "corrupt" and not "real" dialects. Grierson states that for political reasons Bihar has always looked westward even though the language of the area is closer to Bengali and Oriya, being descended from the old Maagadha Apabramsa. It differs from Bengali and Oriya in the pronunciation of the sibilants, striving for a western /s/ rather than an eastern /sh/. Bihari, says Grierson, has three main dialects: Maithilī, Magahī, and Bhojpurī. Each has several subdialects. Maithilī once had a literature, but Magahī (except for the translation of the New Testament in 1818) and Bhojpurī did not. Grierson classified Maithilī and Magahī as one subgroup of Biharī, and Bhojpurī as another. It is not clear why Grierson is so insistent on dialectal status for these three, especially those without literature, when he is not prone to call "dialects" what he considers substandard forms of some languages. He does give morphological peculiarities of some of these Bihari dialects and wherever there is conservatism of older forms from Apabhramsas or wherever, that is clearly good *karma* for that dialect, whereas transitional or mixed lects are not accorded the status of a language group separate from both Hindi and Bengali. Today political groups in Bihar are capitalizing on this to demand language status for their "dialects."[6]

Bengali has, according to Grierson, two regional dialects, East and West. Within Eastern is included Calcutta standard, Southwestern dialect spoken in Midnapore, and Northern Bengali used north of the Ganges; also a western dialect affected by Bihari. Eastern branch includes Dacca dialect (Eastern standard), Rangpur dialect (Raajbangshii), and Southwestern dialect in Chittagong. Grierson also notes the extreme diglossia of the Bengali of his time, which had not yet been alleviated by the developments noted later by Dimock.[7]

Assamese is noted as a language by Grierson, while Beames did not give it this status. Five dialects are noted: Eastern (Standard), Western, Mayaang, Jharwaa, and unspecified. Again, Grierson says that Western Assamese differs little from Eastern, but the only "true dialect" is Mayaang, spoken in Manipur. This could also be a dialect of Bengali, but for various reasons Grierson says it is a dialect of Assamese. Jharwaa, on the other hand, is a "mongrel trade language" in the Garo Hills, being a "pigeon" [sic] mixture of Bengali, Garo, and Assamese. (Grierson, 1903-28:156) Grierson notes that Assamese has been called a dialect of Bengali, and in fact it differs little in grammar from Bengali. However, another test, that of possession of literature, makes Assamese "entitled to claim an independent existence as the speech of an independent nationality, and to have a standard of its own. . ." (Grierson, 1903-28:156)

It should be clear from this that Grierson's theory of what is a language and what is a dialect varies from page to page in the *LSI*. What is used as a criterion for a speech form being classified as a language on one page is used as a criterion for classifying it as a dialect on another page. It also seems that Grierson has an axe to grind in some cases, e.g., in the case of Bihari, where two non-literary forms are called dialects and the whole group is called a language, whereas in the Marathi area Konkani is not accorded this status, although it meets the criterion of having a literature. Note also that Assamese gets to be a language because it has literary history, although it differs little from Bengali.

In concluding our brief discussion on Grierson we would like to make some general remarks concerning the importance of his work. The *LSI* is without doubt the major source of data on which the classification of the modern Indo-Aryan languages has been carried out. It sought to elicit a body of phonological, morphological and lexical information on as wide a variety of spoken vernaculars as possible, as well as to provide sample texts in many of these speech forms. The scope of the problems which it faced was staggering, and its usefulness as a final arbiter of typological questions is therefore somewhat restricted. The *LSI* has been hampered by the lack of a clear cut theory of dialectology through the principles of which decisions might be made concerning the internal divisions of the language family. The work was further hampered by its inability to sort out a confusing array of terms used to represent different speech forms, terms which are offered by interviewed speakers and which often refer to their geographic, religious, or ethnic identity rather than to their purely linguistic identity. Indeed, the *LSI*, as well as language censuses and analyses based on them, often are unable to distinguish among geographic, religious and ethnographic terminology, with frequently no distinction among them noted by the subjects of linguistic and demographic interviews. In spite of these drawbacks, the *LSI* taxonomy of Indo-Aryan languages has served as the starting point for the classification of this family of languages, an endeavor which has not yet moved substantially beyond the beginnings laid down by Grierson. Grierson's full characterizations of the internal relations among the Indo-Aryan languages, expressed as a Stammbaum, and modified slightly by S. K. Chatterji (1926, vol. I, p. 6) is shown below in Figure 28.

3.1.3. Other characterizations of Indo-Aryan

As stated above, the overall typology of Indo-Aryan which has come down to us is essentially that of the *LSI*, a modified thorough analysis of later census statistics. It is, we think, useful to briefly summarize a recent description of the modern Indo-Aryan languages, that of Cardona (1974), in which recent census data has been used. Cardona arranges the Indo-Aryan languages into geographical divisions, with little attempt to construct a formal Stammbaum. Roughly speaking, Cardona considers Indo-Aryan to have eastern, northwestern, western/southwestern, and midlands groups, as well as two others, "Rajasthani" and Bhili, whose vis-á-vis positions are somewhat problematical.

Cardona enumerates Assamese, Bengali and Oriya as constituting the eastern branch of Indo-Aryan. He considers the northwestern group to be composed of Panjabi, Lahnda, Sindhi (this including Kacchi, which according to Cardona, has been claimed by some to be a dialect of Gujarati), the "Pahari" languages, and the Dardic languages (whose positions within the Indo-Iranian family is open to dispute).[8] The Pahari (or "mountain") languages are subdivided into East, West, and Central divisions, the primary example of the first being Nepali, the major examples of the second being Kumaoni and Garwhali, and the last encompassing 62 languages and dialects from the states of Himalchal Pradesh, Jammu and Kashmir, and the Panjab area adjacent to Himalchal Pradesh. Dardic is divided by Cardona into three subgroups: Dard (= East Dardic), Khowar (Central Dardic), and Kafir (West Dardic). The major Dardic language spoken in South Asia is Kashmiri, an East Dardic language. In the western and southwestern divisions Cardona notes Gujarati, Marathi, and Konkani. He also considers Sinhalese, originally an offshoot of a western dialect, to be a member of this group.

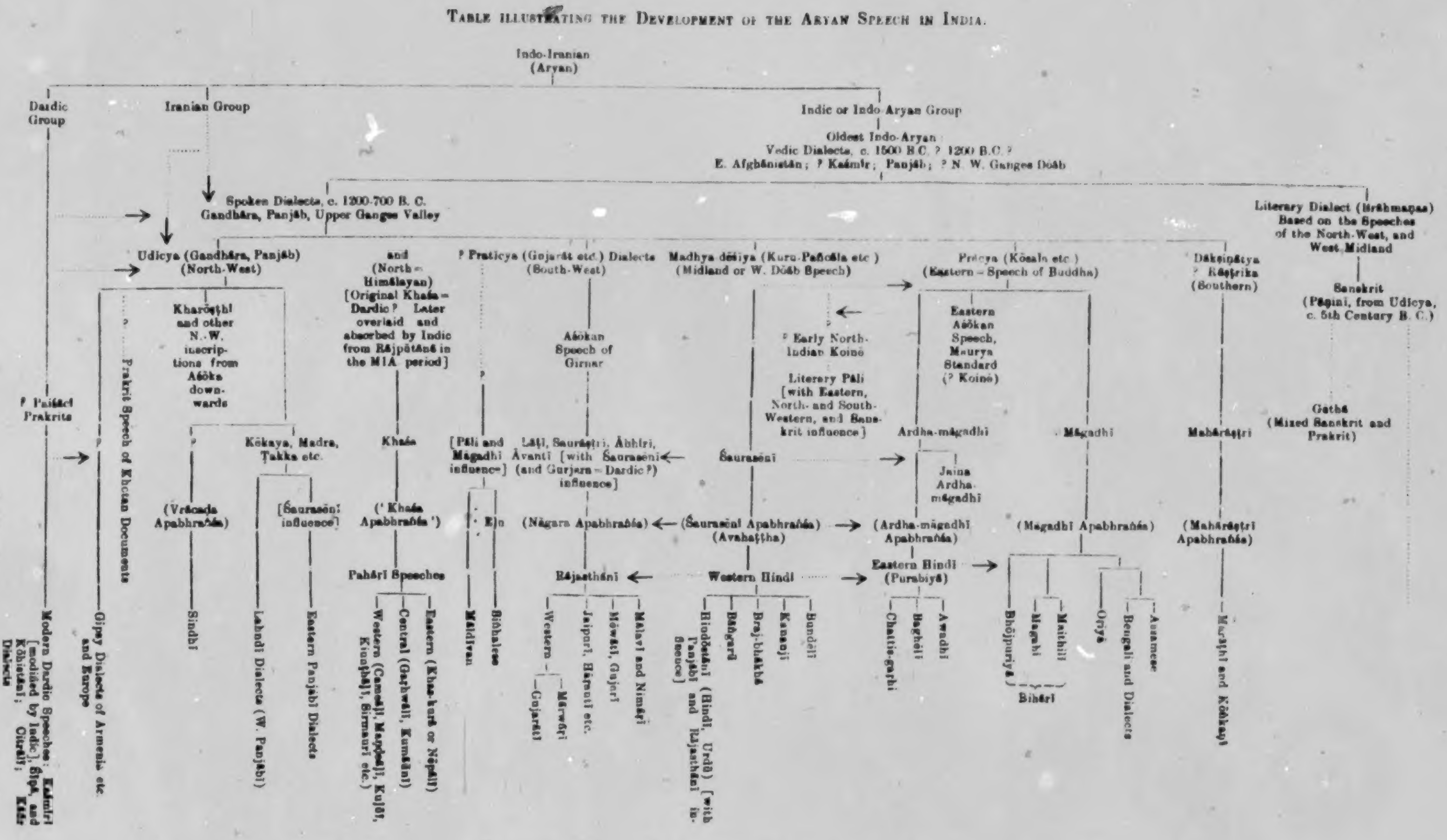


TABLE ILLUSTRATING THE DEVELOPMENT OF THE ARYAN SPEECH IN INDIA.
Indo-Iranian (Aryan)
Dardic Group
Iranian Group
Indic or Indo-Aryan Group
Oldest Indo-Aryan: Vedic Dialects, c. 1500 B.C. ? 1200 B.C. ? E. Afghānistān; ? Kaśmīr; Panjāb; ? N. W. Ganges Dōāb
Spoken Dialects, c. 1200-700 B. C. Gandhāra, Panjāb, Upper Ganges Valley
Literary Dialect (Brāhmaṇas) Based on the Speeches of the North-West, and West Midland
Sanskrit (Pāṇini, from Udīcya, c. 5th Century B. C.)
Gathā (Mixed Sanskrit and Prakrit)
Udīcya (Gandhāra, Panjāb) (North-West)
and (North = Himālayan) [Original Khaśa = Dardic? Later overlaid and absorbed by Indic from Rājpūtānā in the MIA period]
? Pratīcya (Gujarāt etc.) Dialects (South-West)
Madhya dēśiya (Kuru-Pañcāla etc.) (Midland or W. Dōāb Speech)
Prācya (Kōsala etc.) (Eastern – Speech of Buddha)
Dākṣiṇātya ? Rāṣṭrika (Southern)
Kharōṣṭhī and other N.-W. inscriptions from Aśōka downwards
Prakrit Speech of Khotan Documents
? Paiśācī Prakrits
Aśōkan Speech of Girnar
? Early North-Indian Koinē
Literary Pāli [with Eastern, North- and South-Western, and Sanskrit influence]
Eastern Aśōkan Speech, Maurya Standard (? Koinē)
Kēkaya, Madra, Ṭakka etc.
Khaśa
[Pāli and Māgadhī influence]
Lāṭī, Saurāṣṭrī, Ābhīrī, Āvantī [with Śaurasēnī (and Gurjara = Dardic?) influence]
Śaurasēnī
Ardha-māgadhī
Māgadhī
Mahārāṣṭrī
Jaina Ardha-māgadhī
(Vrācaḍa Apabhraṁśa)
[Śaurasēnī influence]
('Khaśa Apabhraṁśa')
Elu
(Nāgara Apabhraṁśa)
(Śaurasēnī Apabhraṁśa) (Avahaṭṭha)
(Ardha-māgadhī Apabhraṁśa)
(Magadhī Apabhraṁśa)
(Mahārāṣṭrī Apabhraṁśa)
Pahāṛī Speeches
Rājasthānī
Western Hindī
Eastern Hindī (Purabiyā)
Modern Dardic Speeches: Kaśmīrī [modified by Indic], Śiṇā, and Kōhistānī; Citrālī; Kāfir Dialects
Gipsy Dialects of Armenia etc. and Europe
Sindhī
Lahndī Dialects (W. Panjābī)
Eastern Panjābī Dialects
Western (Cameālī, Maṇḍeālī, Kulūī, Kiunṭhālī, Sirmaurī etc.)
Central (Gaṛhwālī, Kumāönī)
Eastern (Khas-kurā or Nēpālī)
Māldīvan
Siṅhalese
Western – Gujarātī – Mārwāṛī
Jaipurī, Hāṛautī etc.
Mēwātī, Gujarī
Mālavī and Nimāṛī
Hindōstānī (Hindī, Urdū) [with Panjābī and Rājasthānī influence]
Bāṅgarū
Braj-bhākhā
Kanaujī
Bundēlī
Chattis-gaṛhī
Baghēlī
Awadhī
Bhōjpuriyā
Magahī
Maithilī
Bihārī
Oṛiyā
Bengali and Dialects
Assamese
Marāṭhī and Kōṅkaṇī

Figure 28. The Development of the Aryan Speech in India (Chatterji after Grierson)

Cardona notes a large number of distinct speech forms in the midlands area. In describing the distribution of "Hindi" under its various rubrics he notes that:

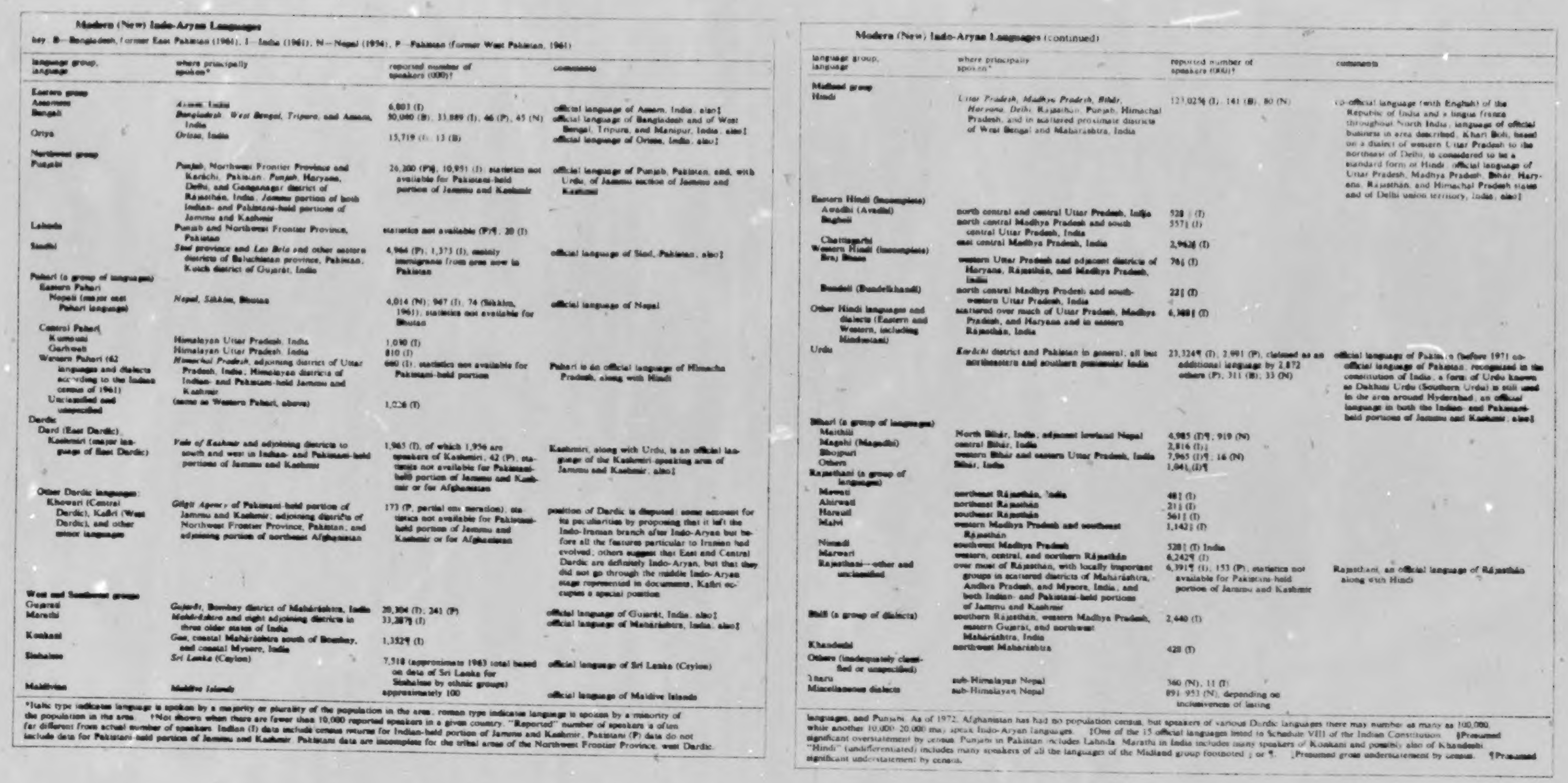
> The midland tract from the borders of Bengal and Orissa to Gujarat and Mahareshtra is a large area where Hindi is the language of official business. The language called *kharī bolī*, considered to be a standard Hindi, is based in a dialect of western Uttar Pradesh to the North-West of Delhi. The term *hindī* (also *hindvī*) is known from as early as the 13th c., when Asmir Khusrau--a minister of the Moghul court--used it. Urdu is also recognized in the constitution of India. . . . The term *zabān-e-urdū* 'language of the imperial camp' came into use about the 17th c. In the south, Urdu was used by Muslim conquerors of the 14th c. and this language, known as Dakhini Urdu ('southern Urdu') is still used in the area about Hyderabad. Structurally and historically Hindi and Urdu are one, though they are now official languages of different countries written in different alphabets. (Cardona, 1974:439)

Cardona also notes the existence of a large number of vernacular "languages"--the midlands area, specifically the Bihari languages (Maithili, Magahi, and Bhojpuri), the Eastern Hindi languages (Avadhi, Bagheli, and Chattisgarhi), Western Hindi (Braj and Bundeli), and Bangru.

A number of languages, Mewati, Ahirwati, Harauti, Malvi, Nimadi, and Marwari, collectively referred to as "Rajasthani" by Grierson, are said by Cardona to represent a shading from "West Hindi" in the east to Sindhi and Gujarati in the west. Another set of languages not enjoying a clear cut position are the Bhili dialects, being concentrated in the area where Madhya Pradesh, Rajasthan and Gujarat meet, as well as in the area where Maharashtra, Madhya Pradesh and Gujarat do so.

A summary of the description of the Indo-Aryan languages as reported by Cardona is given in Figure 29.

In spite of the paucity of attempts since Grierson's time to arrive at overall taxonomies of the Indo-Aryan languages, linguistic studies have been carried out of all of the major languages as well as of substantial numbers of their dialects. It would clearly be impossible to review all of these studies here. For a summary of the work on each of these major languages the interested reader is referred to the articles in *CTL5*.



Modern (New) Indo-Aryan Languages

key: B—Bangladesh, former East Pakistan (1961); I—India (1961); N—Nepal (1954); P—Pakistan (former West Pakistan, 1961)

language group, language	where principally spoken*	reported number of speakers (000)†	comments
Eastern group			
Assamese	*Assam*, India	6,803 (I)	official language of Assam, India, also‡
Bengali	*Bangladesh*, *West Bengal*, *Tripura*, and Assam, India	50,080 (B), 33,889 (I), 46 (P), 45 (N)	official language of Bangladesh and of West Bengal, Tripura, and Manipur, India, also‡
Oriya	*Orissa*, India	15,719 (I), 13 (B)	official language of Orissa, India, also‡
Northwest group			
Punjabi	*Punjab*, Northwest Frontier Province and Karachi, Pakistan; *Punjab*, *Haryana*, Delhi, and Ganganagar district of Rājasthān, India; *Jammu* portion of both Indian- and Pakistani-held portions of Jammu and Kashmir	26,300 (P)§, 10,951 (I); statistics not available for Pakistani-held portion of Jammu and Kashmir	official language of Punjab, Pakistan, and, with Urdu, of Jammu section of Jammu and Kashmir
Lahnda	Punjab and Northwest Frontier Province, Pakistan	statistics not available (P)§, 20 (I)	
Sindhi	*Sind* province and *Las Bela* and other eastern districts of Baluchistan province, Pakistan; Kutch district of Gujarāt, India	4,964 (P), 1,373 (I), mainly immigrants from area now in Pakistan	official language of Sind, Pakistan, also‡
Pahari (a group of languages)			
Eastern Pahari			
Nepali (major east Pahari language)	*Nepal*, *Sikkim*, Bhutan	4,014 (N); 947 (I), 74 (Sikkim, 1961); statistics not available for Bhutan	official language of Nepal
Central Pahari			
Kumauni	Himalayan Uttar Pradesh, India	1,090 (I)	
Garhwali	Himalayan Uttar Pradesh, India	810 (I)	
Western Pahari (62 languages and dialects according to the Indian census of 1961)	*Himachal Pradesh*, adjoining district of Uttar Pradesh, India; Himalayan districts of Indian- and Pakistani-held Jammu and Kashmir	660 (I); statistics not available for Pakistani-held portion	Pahari is an official language of Himachal Pradesh, along with Hindi
Unclassified and unspecified	(same as Western Pahari, above)	1,026 (I)	
Dardic			
Dard (East Dardic)			
Kashmiri (major language of East Dardic)	*Vale of Kashmir* and adjoining districts to south and west in Indian- and Pakistani-held portions of Jammu and Kashmir	1,965 (I), of which 1,956 are speakers of Kashmiri, 42 (P); statistics not available for Pakistani-held portion of Jammu and Kashmir or for Afghanistan	Kashmiri, along with Urdu, is an official language of the Kashmiri-speaking area of Jammu and Kashmir; also‡
Other Dardic languages: Khowari (Central Dardic), Kafiri (West Dardic), and other minor languages	*Gilgit Agency* of Pakistani-held portion of Jammu and Kashmir; adjoining districts of Northwest Frontier Province, Pakistan; and adjoining portion of northeast Afghanistan	173 (P, partial enumeration); statistics not available for Pakistani-held portion of Jammu and Kashmir or for Afghanistan	position of Dardic is disputed: some account for its peculiarities by proposing that it left the Indo-Iranian branch after Indo-Aryan but before all the features particular to Iranian had evolved; others suggest that East and Central Dardic are definitely Indo-Aryan, but that they did not go through the middle Indo-Aryan stage represented in documents; Kafiri occupies a special position
West and Southwest groups			
Gujarati	*Gujarāt*, Bombay district of Mahārāshtra, India	20,304 (I); 241 (P)	official language of Gujarāt, India, also‡
Marathi	*Mahārāshtra* and eight adjoining districts in three older states of India	33,287¶ (I)	official language of Mahārāshtra, India, also‡
Konkani	Goa, coastal Mahārāshtra south of Bombay, and coastal Mysore, India	1,352¶ (I)	
Sinhalese	*Sri Lanka* (Ceylon)	7,510 (approximate 1963 total based on data of Sri Lanka for Sinhalese by ethnic groups)	official language of Sri Lanka (Ceylon)
Maldivian	*Maldive Islands*	approximately 100	official language of Maldive Islands

*Italic type indicates language is spoken by a majority or plurality of the population in the area; roman type indicates language is spoken by a minority of the population in the area. †Not shown when there are fewer than 10,000 reported speakers in a given country. "Reported" number of speakers is often far different from actual number of speakers. Indian (I) data include census returns for Indian-held portion of Jammu and Kashmir. Pakistani (P) data do not include data for Pakistani-held portion of Jammu and Kashmir. Pakistani data are incomplete for the tribal areas of the Northwest Frontier Province, west Dardic languages, and Punjabi. As of 1972, Afghanistan has had no population census, but speakers of various Dardic languages there may number as many as 100,000, while another 10,000–20,000 may speak Indo-Aryan languages. ‡One of the 15 official languages listed in Schedule VIII of the Indian Constitution. §Presumed significant overstatement by census. Punjabi in Pakistan includes Lahnda. Marathi in India includes many speakers of Konkani and possibly also of Khandeshi. "Hindi" (undifferentiated) includes many speakers of all the languages of the Midland group footnoted ‖ or ¶. ‖Presumed gross understatement by census. ¶Presumed significant understatement by census.

Modern (New) Indo-Aryan Languages (continued)

language group, language	where principally spoken*	reported number of speakers (000)†	comments
Midland group			
Hindi	*Uttar Pradesh*, *Madhya Pradesh*, *Bihār*, *Haryana*, *Delhi*, Rajasthan, Punjab, Himachal Pradesh, and in scattered proximate districts of West Bengal and Mahārāshtra, India	123,025§ (I), 141 (B), 80 (N)	co-official language (with English) of the Republic of India and a lingua franca throughout North India; language of official business in area described; Khari Boli, based on a dialect of western Uttar Pradesh to the northeast of Delhi, is considered to be a standard form of Hindi; official language of Uttar Pradesh, Madhya Pradesh, Bihar, Haryana, Rajasthān, and Himachal Pradesh states and of Delhi union territory, India; also‡
Eastern Hindi (incomplete)			
Awadhi (Avadhi)	north central and central Uttar Pradesh, India	528‖ (I)	
Bagheli	north central Madhya Pradesh and south central Uttar Pradesh, India	557‖ (I)	
Chhattisgarhi	east central Madhya Pradesh, India	2,962‖ (I)	
Western Hindi (incomplete)			
Braj Bhasa	western Uttar Pradesh and adjacent districts of Haryana, Rājasthān, and Madhya Pradesh, India	70‖ (I)	
Bundeli (Bundelkhandi)	north central Madhya Pradesh and southwestern Uttar Pradesh, India	22‖ (I)	
Other Hindi languages and dialects (Eastern and Western, including Hindustani)	scattered over much of Uttar Pradesh, Madhya Pradesh, and Haryana and in eastern Rājasthān, India	6,388‖ (I)	
Urdu	*Karāchi* district and Pakistan in general; all but northeastern and southern peninsular India	23,324¶ (I), 2,991 (P), claimed as an additional language by 2,872 others (P); 311 (B); 33 (N)	official language of Pakistan (before 1971 co-official language of Pakistan); recognized in the constitution of India; a form of Urdu known as Dakhini Urdu (Southern Urdu) is still used in the area around Hyderabad; an official language in both the Indian- and Pakistani-held portions of Jammu and Kashmir; also‡
Bihari (a group of languages)			
Maithili	North Bihār, India, adjacent lowland Nepal	4,985 (I)¶; 919 (N)	
Magahi (Magadhi)	central Bihār, India	2,816 (I)‖	
Bhojpuri	western Bihār and eastern Uttar Pradesh, India	7,965 (I)¶; 16 (N)	
Others	Bihār, India	1,041 (I)¶	
Rajasthani (a group of languages)			
Mewati	northeast Rājasthān, India	48‖ (I)	
Ahirwati	northeast Rajasthan	21‖ (I)	
Harauti	southeast Rājasthān	561‖ (I)	
Malvi	western Madhya Pradesh and southeast Rājasthān	1,142‖ (I)	
Nimadi	southwest Madhya Pradesh	528‖ (I) India	
Marwari	western, central, and northern Rājasthān	6,242¶ (I)	
Rajasthani—other and unclassified	over most of Rājasthān, with locally important groups in scattered districts of Mahārāshtra, Andhra Pradesh, and Mysore, India, and both Indian- and Pakistani-held portions of Jammu and Kashmir	6,391¶ (I); 153 (P), statistics not available for Pakistani-held portion of Jammu and Kashmir	Rajasthani, an official language of Rājasthān along with Hindi
Bhili (a group of dialects)	southern Rājasthān, western Madhya Pradesh, eastern Gujarat, and northwest Mahārāshtra, India	2,440 (I)	
Khandeshi	northwest Mahārāshtra	420 (I)	
Others (inadequately classified or unspecified)			
Tharu	sub-Himalayan Nepal	360 (N); 11 (I)	
Miscellaneous dialects	sub-Himalayan Nepal	891–953 (N), depending on inclusiveness of listing	

Figure 29. Cardona Summary

3.2. *Classification of the Dravidian languages*

Dialectological studies in the Dravidian area are somewhat clearer and perhaps more theoretically interesting than in the Indo-Aryan area, perhaps because questions of language classification in the Dravidian area are not as cloudy as in north India, and there has been general agreement in the South that a given speech variety is a dialect in some sense of one of the major languages, and not a dialect of some other major language. Nevertheless, there do exist cases where disagreement about subgrouping has reigned, and this includes both speech varieties which have been known for some time as well as some which have only recently been classified.

The Dravidian languages have been believed to be a separate family since the time of Caldwell, and indirectly, since Ellis before him.[9] Caldwell's scheme of classification did not include all the presently known languages, since in his time many had not been enumerated or described, although grammars and dictionaries of many of the larger languages had already appeared by his time.

Caldwell presented no tree diagrams or other data to indicate how he viewed the subgrouping of Dravidian languages. Grierson, however, basing his work on Caldwell, presents the Stammbaum shown in Figure 30. The classification in Figure 30 displays a recognizable South Dravidian group, with Tamil and Malayalam closely related, Tulu and Kodagu less so, and a separate subbranch with Kanarese flanked by Toda and Kota. Until the work of Emeneau in the 30's and 40's of this century, the idea that Toda and Kota were either dialects of or closely related to Kannada was generally accepted. Another recognizable group, but not named as such, is a Central group with Kurukh and Malto closely related, flanked by Gondi, Kui, Kolami, etc., and finally Telugu figuring as a somewhat distant relative of the other Central Dravidian languages. Finally, in complete isolation, is Brahui. Later schemes have Kurukh and Malto removed from the Central branch and placed with the other northern language, Brahui. Telugu also moves to a closer position with relation to South Dravidian, and according to some, is a South Dravidian language with affinities to Central.[10]

Since the *LSI* was not carried out in those areas of South India where many Dravidian languages are spoken, but merely incorporated dialects and languages from the Dravidian area which happened to be spoken in the area

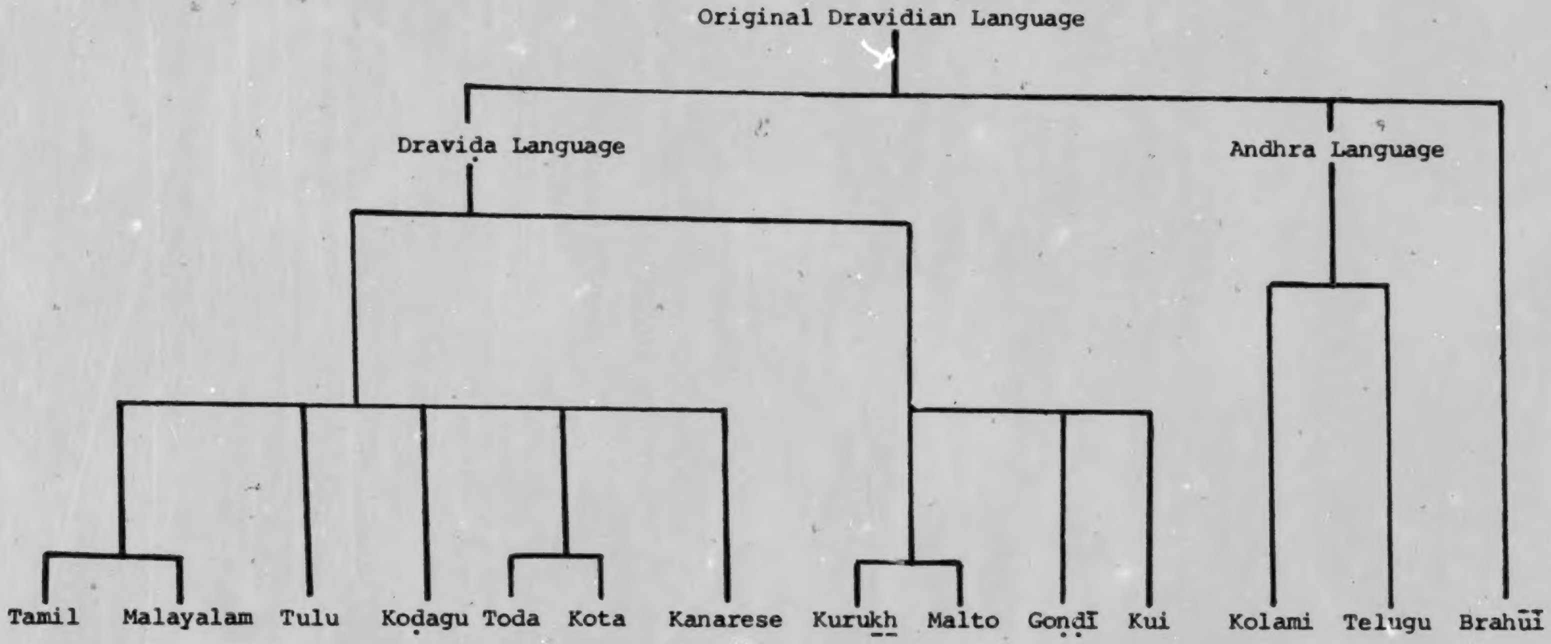

Original Dravidian Language
Draviḍa Language
Andhra Language
Tamil
Malayalam
Tulu
Koḍagu
Toda
Kota
Kanarese
Kuruḵẖ
Malto
Goṇḍī
Kui
Kolami
Telugu
Brahūī

Figure 30. Classification of the Dravidian Languages (Grierson after Caldwell)

covered by the survey, only a few Dravidian dialects are cited. Many of these would probably today not be considered dialects of the languages to which they are attributed. For example, the dialects of Tamil listed in the Tamil section are today either considered by linguists as separate languages (cf. Diffloth, 1968; Zvelebil, 1973) or are considered dialects of other languages, e.g. of Kannada.

Grierson's classification of Dravidian languages is carried out on essentially morphological grounds. Thus he groups Tamil and Kannada together because they both preserve a "regular feminine gender" (*LSI*, vol. 4, 284), although he notes that Kannada and Telugu both possess a present participle, unlike Tamil. On the basis of verb tense formation he classes Kurukh̲ and Malto with "the same dialect as that which became the common origin of Tamil and Kanarese." (*LSI*, vol. 4, 284) After the time of Caldwell (and Grierson's repetition of Caldwell's scheme) no major revisions of the Dravidian family tree structure were proposed until the 1950's, when new work on various languages necessitated inclusion of new languages and revision of the earlier scheme. (Some discussion of various possibilities, mainly on historical grounds, occurred in the works of L. V. Ramaswami Aiyar, E. H. Tuttle, and T. Burrow, but no specific subgroupings were proposed). With the publication of their *Parji Language* Burrow and Bhattacharya state the close relationship between Parji, Ollari, Gadaba, Kolami, Naiki, Gondi-Konda, and Kui-Kuvi, with special subgroupings within those. (Burrow and Bhattacharya 1953:xi)

Emeneau (1955) posited a close connection between Kolami and Naiki, Parji, and Ollari, grouping these languages together into what he called the Kolami-Parji subgroup. He also proposed some other tentative subgroupings within Central Dravidian. In fact, Emeneau proposes that Kolami and Naiki are probably dialects of the same language, with probable great mutual intelligibility. (Emeneau, 1955:141-2). However, since only word lists were available at that time for some of the languages, no morphological comparisons could be made.

In 1961 Krishnamurti (Krishnamurti 1961:236-74) established that Telugu, long thought to be a South Dravidian language, was actually a Central language with strong affinities to the Kui-Kuvi-Konda subgroup, but which had had close geographic contact with the South Dravidian languages for a long time.

In 1962 Emeneau (Emeneau, 1962b:62-70) proposed that Brahui and Kurukh-Malto are probably a subfamily, due to certain phonological isoglosses and some common retentions and shared innovations; but he states that additional evidence is needed to show conclusively that Brahui is closer to Kurukh-Malto, than to some other Dravidian language, or whether some other kind of branching is to be proposed.

Despite all these proposals for subgrouping within the Dravidian family, no Stammbaum diagrams are proposed to summarize all the new evidence until Andronov's work in the early and mid-1960's.[11] Figure 31 was constructed to show the overall classification of the Dravidian languages according to Andronov. As is obvious, Kurukh and Malto are now grouped with Brahui, forming what is now generally called a North Dravidian group; Tulu is removed from its central place among the South Dravidian languages and placed at one side with Telugu, while Toda and Kota are moved to a place closer to Tamil-Malayalam. Additional Central Dravidian languages are included with that group, and subgrouped in pairs of pairs. Andronov is not explicit in his criteria for grouping the languages the way he does, but subsequent work seems to substantiate much of his scheme. In a later work (Andronov, 1970) a minor revision of the Central Dravidian portion of the tree is proposed, and some other languages are moved higher on the tree to yield the schema shown in Figure 32.

This scheme removed Telugu and Tulu from South Dravidian per se and proposes that they both separated from South Dravidian at the same time as the other major subgroups, which now include North Dravidian, Gondi-Kondi-Kui-Kuvi, and Kolami-Naiki-Parji-Gadaba (these last two forming in the 1963/4 scheme a central group). As Andronov notes, "genetic affiliations inside the Dravidian linguistic system have not yet been completely clarified!"[12] He calls these new subgroupings *southern* (Tamil to Kannada), *southwestern* (Tulu), *southeastern* (Telugu), *central* (Kolami, Naiki, Parji, Gadaba), *Gondwana* (Gondi, Konda, Kui, Kuvi), *northeastern* (Kurukh and Malto) and *northwestern* (Brahui). The splitting of the northern group into two groups is not reflected in his diagram. This scheme is based on Andronov's own lexicostatistic work (Andronov 1964a) on the disintegration of the Dravidian languages.

As we have noted, the earliest reference to a North Dravidian group of languages consisting of Brahui on the one hand and Kurukh-Malto on the other is in Emeneau, 1962b. Emeneau gives no evidence that he believes

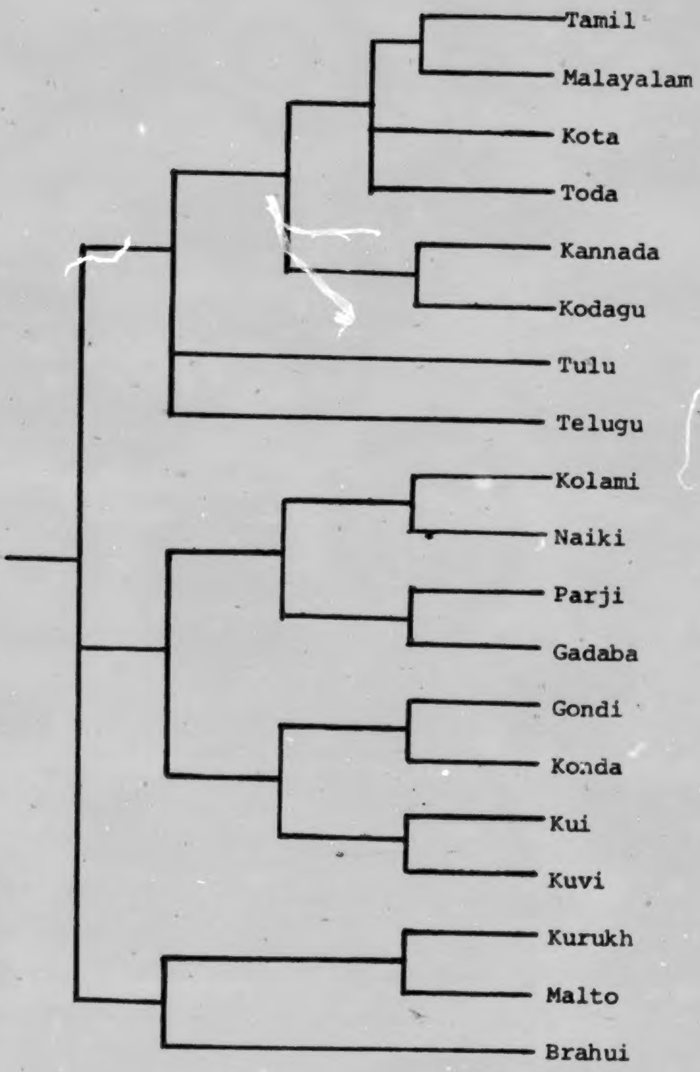

Tamil
Malayalam
Kota
Toda
Kannada
Kodagu
Tulu
Telugu
Kolami
Naiki
Parji
Gadaba
Gondi
Konda
Kui
Kuvi
Kurukh
Malto
Brahui

Figure 31. Classification of the Dravidian Languages (after Andronov)

b

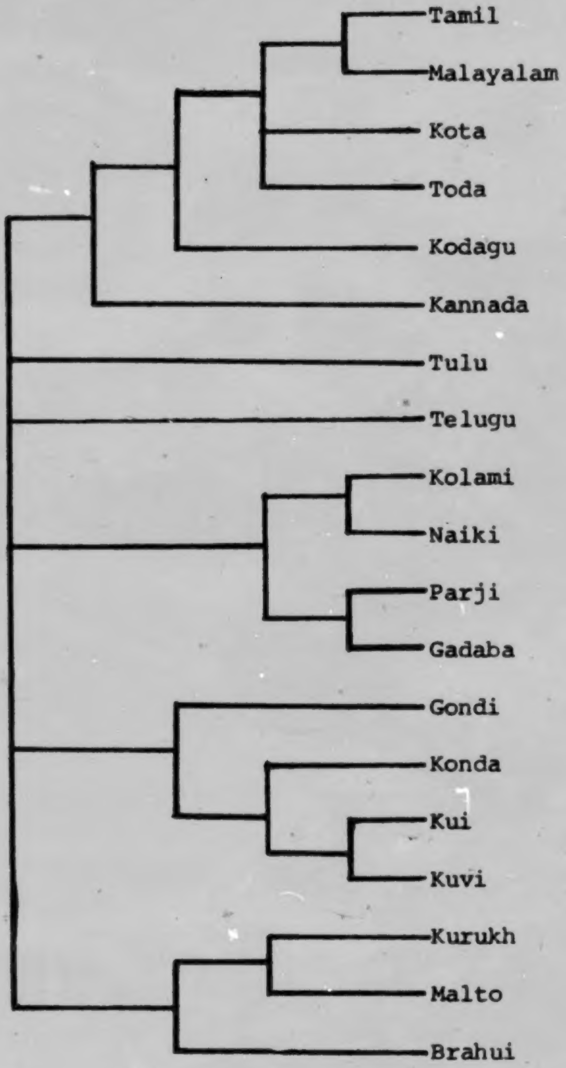

Tamil
Malayalam
Kota
Toda
Kodagu
Kannada
Tulu
Telugu
Kolami
Naiki
Parji
Gadaba
Gondi
Konda
Kui
Kuvi
Kurukh
Malto
Brahui

Figure 32. Classification of the Dravidian Languages (Andronov 2)

that the two subgroups are to be traced directly to Proto-Dravidian rather than Proto-north Dravidian.

A Central Dravidian group of languages consisting of two subgroups is first proposed by Burrow and Bhattacharya (1953). Its position has been modified in subsequent publications. P. S. Subrahmanyam (1969) proposes a Proto-Central Dravidian tree structure which includes two newly discovered languages, Pengo and Manda, but which preserves the bifurcation into a Kolami-Naiki-Parji-Gadaba (KNPG) subgroup and a Proto-Telugu-Kui one (Figure 33 below). Subrahmanyam cites Krishnamurti (1961) as containing the best morphological evidence for including Telugu with Central Dravidian, even though the language has strong ties to South Dravidian phonologically, participating in the loss of initial /c/ and the palatalization of initial /k/.

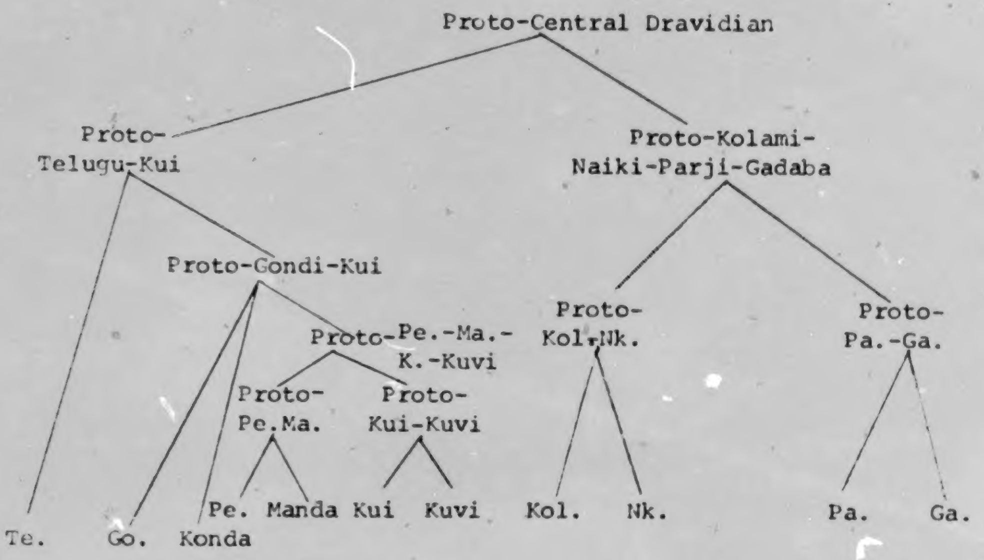

Proto-Central Dravidian
Proto-Telugu-Kui
Proto-Kolami-Naiki-Parji-Gadaba
Proto-Gondi-Kui
Proto-Pe.-Ma.-K.-Kuvi
Proto-Pe.Ma.
Proto-Kui-Kuvi
Proto-Kol.-Nk.
Proto-Pa.-Ga.
Te.
Go.
Konda
Pe.
Manda
Kui
Kuvi
Kol.
Nk.
Pa.
Ga.

Figure 33. The Central Dravidian Languages (after Subrahmanyam)

The diagram given here differs slightly from a similar one in Subrahmanyam, 1968, the latter not including Pengo and Manda. The diagram has also been altered somewhat from the form in which it was presented by Subrahmanyam, and incorporates new information from Burrow, specifically, that Pengo and Manda are closely related to one another and, as a subgroup, are more related to Kui-Kuvi than to Proto-Gondi-Kui, this in spite of close contact with Konda (= Kubi).

Subrahmanyam's diagram has Pengo and Manda branching off from Proto-Gondi-Kui, parallel to Gondi and Konda. Manda, discovered by Burrow and Bhattacharya while working on Pengo, has not been closely studied, but from the initial research of these linguists a close relationship with Pengo seems likely.

The primary source for classifying the south Dravidian languages is Emeneau, 1967. His system is based on morphological considerations, primarily the shape of tense morphemes in the various languages, and is reasonably authoritative except for the exclusion of Tulu and some other dialects discovered or worked on more recently than 1967 (i.e., Irula, Kurumba/Kuruba, and Koraga). We will discuss their status below. Emeneau mentions that Tulu is introduced in one of his diagrams as a language of interaction with Kannada and Kodagu, but no claims are made for it (or Telugu) in terms of genetic relationships with PSDr. The overall schema for the South Dravidian languages given in Emeneau, 1967 is reproduced below in Figure 34.

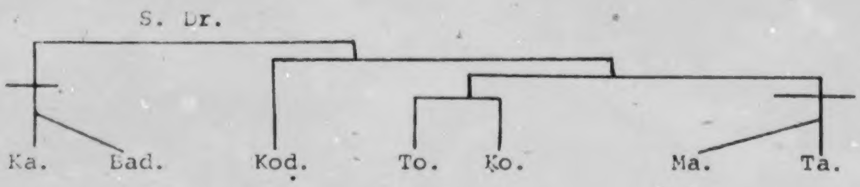


Figure 34. The South Dravidian languages (from Emeneau, 1967:370) Ka= Kannada, Bad= Badaga, Kod= Kodagu, To= Toda, Ko= Kota, Ma= Malayalam, Ta= Tamil. For Ka. and Ta. —— indicates approximately the beginning of the literary record.

Another diagram also provided in Emeneau, 1967 (Figure 35), is designed to incorporate the degree of interaction among the South Dravidian languages, as well as to show their genetic relationships. This diagram is supposed to represent the spatial relationships between the languages as well as the genetic relationships, with solid lines representing historical relationships and the dotted ones showing the interaction of languages with one another. As such it is a departure from previous stemmata diagrams in that social factors which have influenced the history of the languages are sketched. It is therefore essentially a Stammbaum diagram with an extra dimension added to incorporate

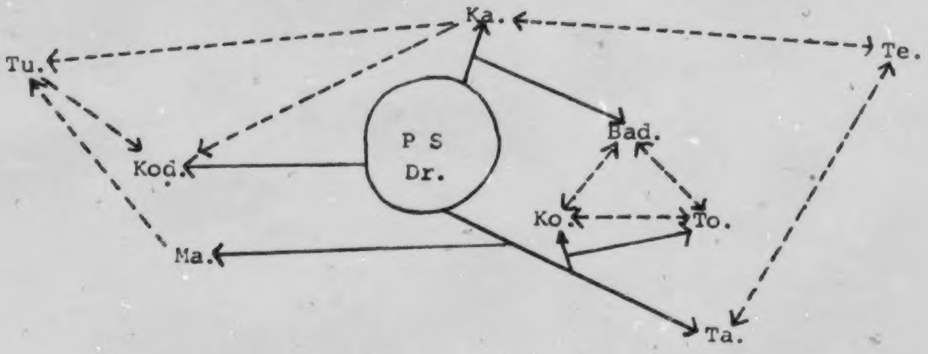


Figure 35. A three-dimensional model of South Dravidian (from Emeneau, 1967:371).

nongenetic influences on the development of the language family.

A similar attempt to incorporate areas of influence among the South Dravidian languages into a purely genetic framework is provided by Zvelebil in his *Comparative Dravidian Phonology* (The Hague: Mouton, 1970). His results are reproduced below in Figure 36.

An additional description of the relations among the South Dravidian languages is to be found in Bh. Krishnamurti's article in *CTL5* (Krishnamurti, 1969a). In this article Krishnamurti summarizes earlier studies on the subgroupings within the language family, and proposes a number of Stammbaums. His tree structure for South Dravidian does not include Pengo-Manda (which was not described until after the completion of Krishnamurti's paper), but includes a node for Ollari, which many writers have felt to be merely a dialect of Gadaba. Krishnamurti's basis for classification is both phonological and morphological. It is interesting to note that while Malayalam is called an off-shoot of Tamil, and given a node as a language, Badaga, called an offshoot of Kannada, is not given a node by Krishnamurti. Krishnamurti's classification of South Dravidian is given in Figure 37. In the same work Krishnamurti (1969a) also presents Stammbaums for the North and Central Dravidian languages. These are reproduced below in Figures 38 and 39.

In addition to treating the South Dravidian languages, Zvelebil in his *Comparative Dravidian Phonology* attempts to construct a Stammbaum for the entire Dravidian family of languages. His scheme of classification gives a place to several languages not hitherto included in tree diagrams, to wit, Irula, Badaga, Savara (close to Telugu), various

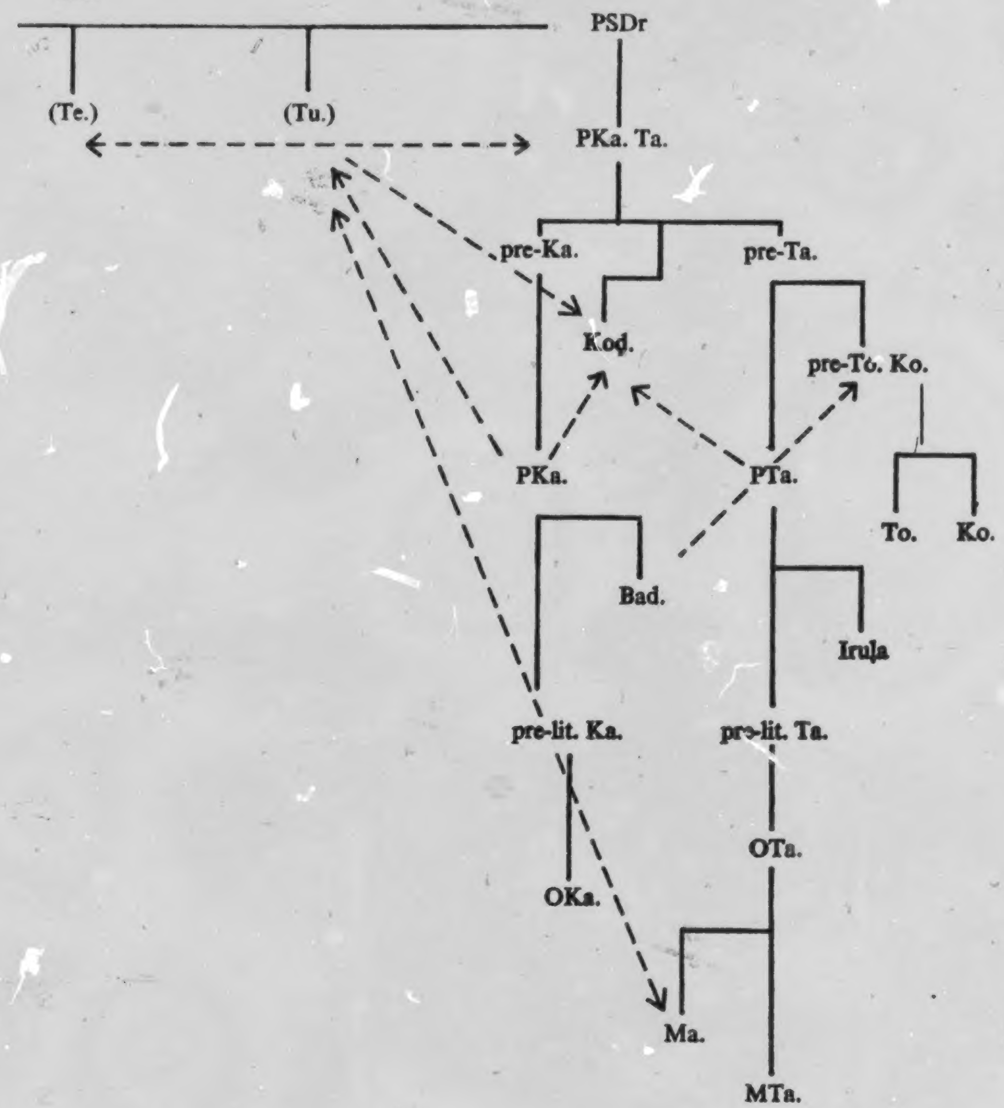

PSDr
(Te.)
(Tu.)
PKa. Ta.
pre-Ka.
pre-Ta.
Koḍ.
pre-To. Ko.
PKa.
PTa.
To.
Ko.
Bad.
Iruḷa
pre-lit. Ka.
pre-lit. Ta.
OTa.
OKa.
Ma.
MTa.

Figure 36. The South Dravidian Languages (Zvelebil)

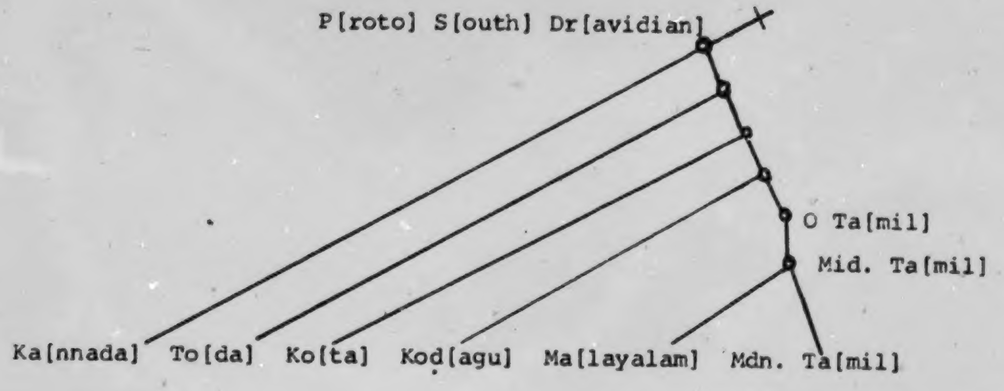
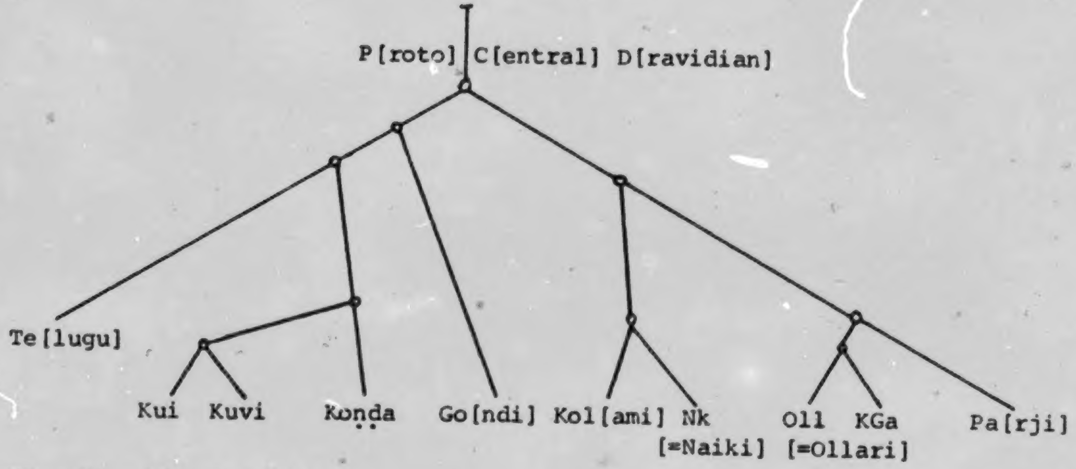

P[roto] S[outh] Dr[avidian]
O Ta[mil]
Mid. Ta[mil]
Ka[nnada]
To[da]
Ko[ta]
Koḍ[agu]
Ma[layalam]
Mdn. Ta[mil]

Figure 37. The South Dravidian Languages (from Krishnamurti, 1969a:327)


P[roto] C[entral] D[ravidian]
Te[lugu]
Kui
Kuvi
Koṇḍa
Go[ndi]
Kol[ami]
Nk
[=Naiki]
Oll
[=Ollari]
KGa
Pa[rji]

KGa= K[ondekor] Ga[daba]

Figure 38. The Central Dravidian languages (from Krishnamurti, 1969a:327)

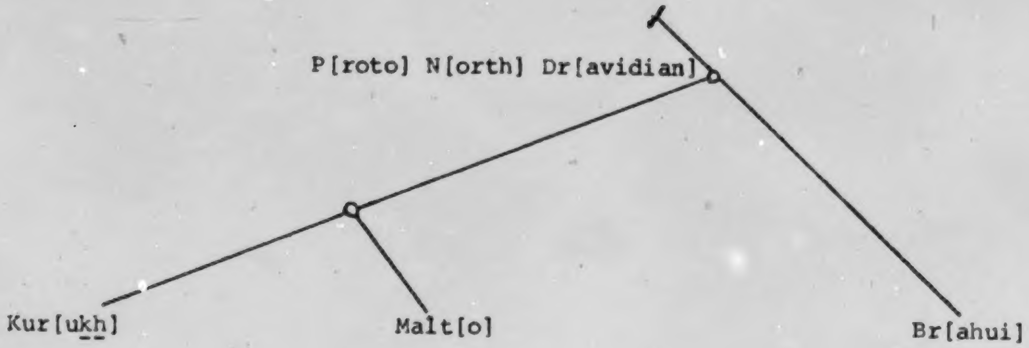

P[roto] N[orth] Dr[avidian]
Kur[ukh]
Malt[o]
Br[ahui]

Figure 39. The North Dravidian languages (from Krishnamurti, 1969a:327)

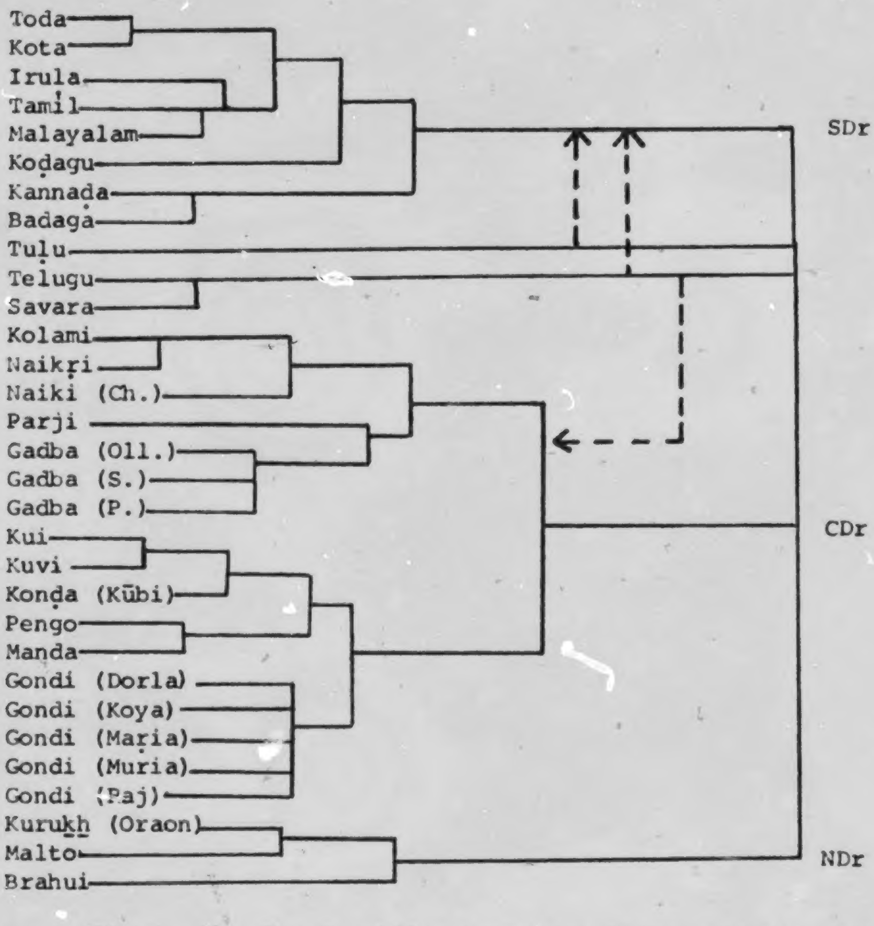

Toda
Kota
Irula
Tamil
Malayalam
Kodagu
Kannada
Badaga
Tulu
Telugu
Savara
Kolami
Naikri
Naiki (Ch.)
Parji
Gadba (Oll.)
Gadba (S.)
Gadba (P.)
Kui
Kuvi
Konda (Kūbi)
Pengo
Manda
Gondi (Dorla)
Gondi (Koya)
Gondi (Maria)
Gondi (Muria)
Gondi (Raj)
Kurukh (Oraon)
Malto
Brahui
SDr
CDr
NDr

Figure 40. The Dravidian Languages (From Zvelebil, *Comparative Dravidian Phonology*, p. 13)

Gadaba dialects (spelled Gadba), Pengo and Manda, and many Gondi dialects. His scheme is reproduced in Figure 40. Zvelebil does not give immediately adjacent to this tree diagram any justification for its nodes; one must presumably look elsewhere for the criteria used in determining them. This chart is interesting; it assigns a place to Tulu and Telugu intermediate between SDr and CDr but notes contactual relationships between Tulu and South Dravidian and between Telugu and both South Dravidian and Central Dravidian. Zvelebil enumerates 31 languages in all, and also mentions others whose place is not yet certain, e.g. Kaikāḍi, Kāḍar, Yerukala and others.

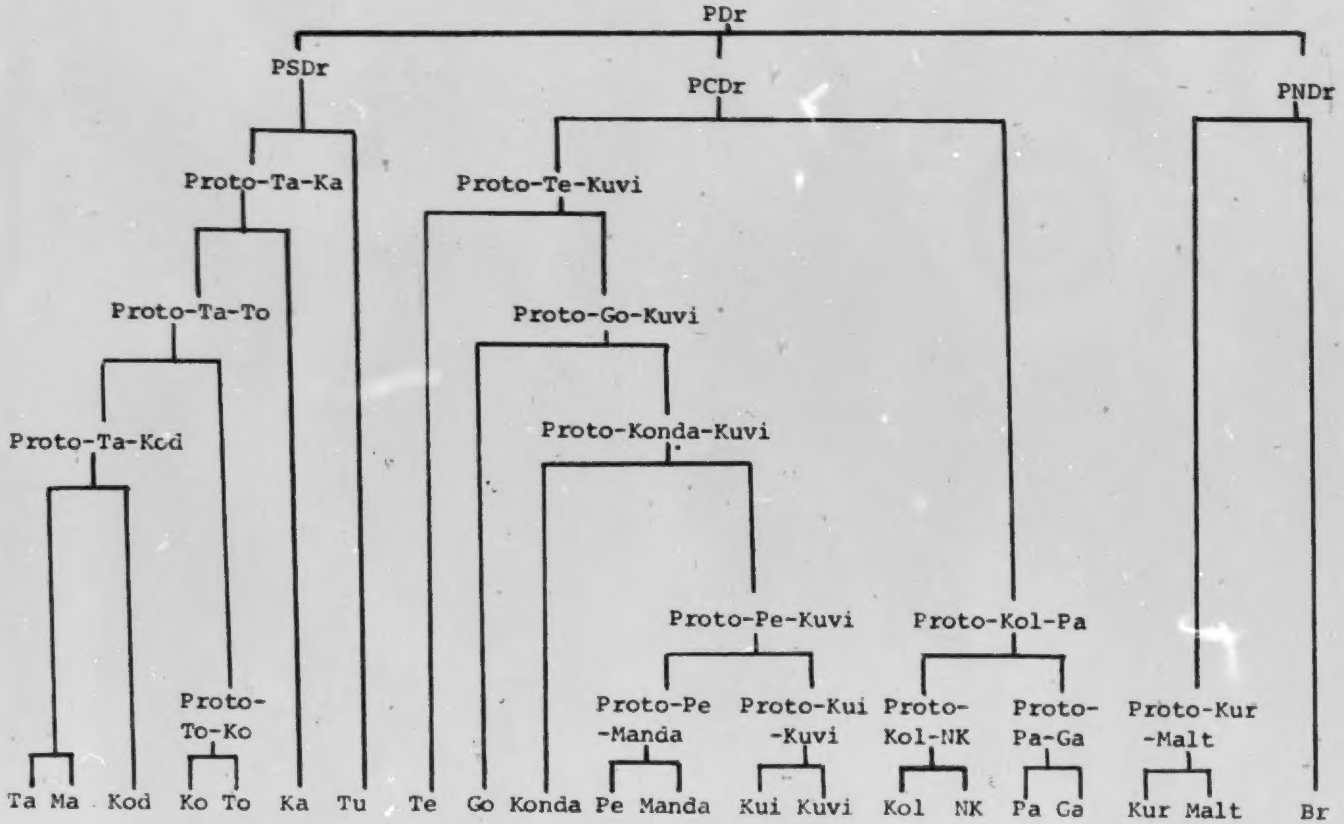
Probably the most recent typology of the entire Dravidian family of languages is to be found in P. S. Subrahmanyam's *Dravidian Verb Morphology* (1971). This work contains a chart (Figure 41) which incorporates Pengo and Manda and also makes slight alterations in the earlier places assigned to Gondi and Konda. The chart demonstrates a closer relationship between Kodagu and Tamil-Malayalam than shown in earlier work. Evidence given for the differentiation among Telugu, Gondi, and Konda is primarily morphological, although some phonological evidence is also given. The evidence for a separate Tamil-Kodagu subgroup, remaining after the branching of Toda-Kota is tentative; Subrahmanyam attributes the claim to Bh. Krishnamurti (1969a) who points out that these languages retain "derivative a in verbs (corresponding to class VII verbs of Tamil Tamil. . .)." Subrahmanyam finds this criterion for the establishment of a common Tamil-Kodagu group less convincing than the shared innovation of use of the plural -*kaḷ* with certain pronoun stems.

The Stammbaum in Figure 41 fails to incorporate a number of recent developments in Dravidian subgrouping, as well as to even mention some named speech forms which have had claims made for them as separate languages, such as Badaga, Koya, Ollari, Kuruba/Kurumba, Irula, and Koraga.

We will now discuss the status of these last named speech forms, and attempt to incorporate them into a final scheme, or relegate them to the status of dialects, and discuss them in succeeding sections of this chapter.

Badaga. Badaga seems to get support from Emeneau (1967) as an early offshoot of Kannada, participating in the Kannada change of *#p- to #h-, etc. But Badaga differs from Kannada in its causative formation and in

Figure 41. Stammbaum of the Dravidian languages (from Subrahmanyam, 1971:531)


PDr
PSDr
PCDr
PNDr
Proto-Ta-Ka
Proto-Te-Kuvi
Proto-Ta-To
Proto-Go-Kuvi
Proto-Ta-Kod
Proto-Konda-Kuvi
Proto-Pe-Kuvi
Proto-Kol-Pa
Proto-To-Ko
Proto-Pe-Manda
Proto-Kui-Kuvi
Proto-Kol-NK
Proto-Pa-Ga
Proto-Kur-Malt
Ta
Ma
Kod
Ko
To
Ka
Tu
Te
Go
Konda
Pe
Manda
Kui
Kuvi
Kol
NK
Pa
Ga
Kur
Malt
Br

a number of other points. However, as Emeneau points out, Kannada exhibits a number of striking deviations from Kannada-like features, and seems to have developed some morphological traits, like the inclusive-exclusive distinction in pronouns, as well as a tense-marker distinction between transitive and intransitive pairs, under pressure from the Nilgiris languages or other South Dravidian languages. Emeneau stops short of declaring Badaga to be a language separate from Kannada rather than a dialect of it, in the absence of a detailed description of it and other dialects of Kannada. Probably we have here an example of dialectological theory which bows to the tradition of the great tradition of Literary Kannada as being the major language, with unlettered Badaga a mere dialect, whereas if Kannada and Badaga were both non-literary languages, Badaga would surely come out as a language rather than a dialect. Coupled with Emeneau's scholarly reluctance to declare anything a fact without adequate demonstration, Emeneau's verdict seems to be wait and see. We therefore concur in leaving Badaga as a dialect of Kannada until further evidence shows it not to be such.

Koya. In his monograph on Koya (Tyler, 1969b), Tyler concludes that "since Koya is a Gondi language, it is mutually intelligible with Hill Maria Gondi in Bastar and Sironcha." (Tyler, 1969b:3) He further notes that ". . .the general pattern seems to be for geographically adjacent Koya and Gondi populations to speak different, but mutually intelligible Gondi dialects. Where these populations are geographically non-contiguous, the dialects are not mutually intelligible. This same pattern probably prevails among all Gondi dialects." (Ibid.) The term "Gondi" therefore seems to refer to a chain of dialects in which mutual intelligibility decreases with distance. Koya is one of the names of some of the Gondi dialects (since there are a number of varieties of Koya). We will therefore await, with Tyler, further work on Koya and Gondi for a definitive statement on these speech forms.

Irula. Irula has been the subject of a Ph.D. dissertation by Diffloth (1968) and a monograph by Zvelebil (1973). Diffloth considers Irula a Dravidian language of the Tamil-Malayalam group which is a close relative of Tamil. That it is not a dialect of Tamil or another Dravidian language is shown by its conservative phonology, (lack of palatalization of initial velars before /i/, lack of palatalization in the past tense formation after front vowels), its preservation of three apical stops, its loss of the retroflex liquid /r̤/, and on morphological grounds. Zvelebil (1973) also considers Irula an independent language, rather than a dialect, but notes that "the question whether Irula is a separate (independent) South Dravidian language, or an (archaic) Tamil dialect, cannot, . . . be settled quite definitively at this moment." (Zvelebil, 1973:2). He notes that the status of Irula depends on our definition of language and dialect, and because these two terms are not purely linguistic terms, "neither are the criteria to distinguish them." (Zvelebil, 1973:3) On mutual intelligibility grounds, however, Irula is a language separate from Tamil. Furthermore, Zvelebil regards the overall morphological patterning of Tamil dialects to be too narrow to include the structure of Irula.

On the basis of these two works, therefore, we should consider Irula to be an independent language, but closely related to Tamil.

As for other Nilgiris tribes, Diffloth recapitulates the names of tribes mentioned in anthropological sources, and notes the scanty linguistic data available on them. But some affinities with Irula can be noted from the Gravely materials on Kasuva,[13] while Yerukala-Korava seems to be mutually unintelligible with Irula.[14]

Kuruba/Kurumba. There seem to be a number of speech forms known either as Kuruba or Kurumba. Diffloth notes the presence of Pāl Kurumbas in the Nilgiris but ventures no classification of their speech except to state that it is not the same as Irula. He also cites Beṭṭa Kurumba, Muḷḹu Kurumba, and Jēn Kurumba, as well as some other "new Dravidian languages": Paṇiya (Tamil-Malayalam group) and Sholaga (Kannada group). Beṭṭa Kurumba is declared to be South Dravidian by Diffloth.[15]

Kuruba, a speech form spoken by Beṭṭa Kurubas in Coorg district, is called a language by Upadhyaya (1972) and grouped with Kodagu in South Dravidian. The evidence given for calling this a language and not a dialect of something else is phonological (presence of retroflex vowels, absence of sibilants, change of medial *ḍ* to *jj*) and morphological (accusative, plural, and formative suffixes unlike other languages, etc.). It is grouped with Koḍagu because of lack of similarity with Kannada or Nilgiris languages and because it has retroflex vowels and other phonological similarities to Koḍagu. It is still not clear whether Beṭṭa Kurumba and other varieties of Kurumba are related to Beṭṭa Kuruba or not.

Jeinu Kurubas and Aine Kurubas speak a dialect of Kannada.

Koraga. Koraga, earlier grouped with Tulu as a dialect of Tulu despite many differences, is classified as a language or perhaps even two languages, by D. N. S. Bhat.[16] Apparently the two dialects, Korra and Mu:ḍu, are quite distinct and not mutually intelligible with each other, with Tulu,[17] or with Kannada. In fact, Koraga seems to show some affinities with north Dravidian, e.g. in the gender-number system 3rd person, similar to Kuṛux, the present tense suffix in -n-, the past tense suffix in /-k-/ or /-g-/ and some others. The two dialects seem to have separated before the change of Kannada initial /p/ to /h/ since the Mu:ḍu dialect participated in the Kannada change, but the Korra dialect did not. Also Proto-Dravidian /*r̤/ in Korra is represented by /r/ and in Mu:ḍu by /ḷ/, another change which took place in Kannada in the tenth century (merger of r̤ and ḷ). If Koraga is closely related to North Dravidian, it has clearly been in close geographical contact with Kannada and Tulu respectively for perhaps a millenium.

Dialects of Tamil. In the earliest grammar of Tamil extant, Tolkaappiyam, twelve divisions of the Tamil country, each with its own dialect, are indicated. The differences among them are phonological, lexical, and syntactic. Any deviations from the standard literary dialect were characterized as vulgar and were warned against. In Tolkaappiyanaar's time Kerala was included in Tamil Nadu, as Malayalam had not yet diverged from Tamil. Ceylon dialects, however, are not mentioned in the work, since Tamil speakers had apparently not settled in Ceylon at that time.

T. P. Meenakshisundaran (1965:194-217) enumerates a number of modern Tamil dialects, such as the Ceylon dialects, the Tamil of Malaya, Burma, and South Africa; the Tigalu dialect of Bangalore, the Harijan dialect of Bangalore, the Sanketi dialect in Mysore; the Hebbar and Mandyam Brahman's dialect, the Secunderabad dialect of Brahmans settled in Andhra, and of course the colloquial dialects of Tamil Nadu proper. Few examples are given of the foregoing dialects and how they are distinguished from others, except for one phonological and one lexical example.

Grierson mentions the following dialects of Tamil. Yerukala or Korava (spoken by a wandering tribe), Irula (a caste dialect in the Nilgiris, etc.) and Kasuva (dialect of a jungle tribe), Kaikaaḍī and

Burgaṇḍī (two other dialects spoken by vagrant Gipsy tribes). Despite the statement that only the spoken dialect of Tamil will be considered in the survey, the specimen of Tamil given is literary. Specimens of the other dialects mentioned are given in what one assumes is the ordinary pronunciation of those speech forms. Many of them have affinities with Kanarese, as Grierson notes, and as is obvious to the present day reader (for example, initial /h/ in forms where Tamil has /p/.)

A large number of dialect studies by Zvelebil (1959, 1960, 1963a, b, 1964 and 1966) lay out a pattern of nine main types of Tamil: Literary Standard, Colloquial Standard, Brahmin speech (Aiyar and Aiyangar dialects), and the regional dialects: Northern Tamil, Western, Eastern, Southern, Ceylon dialects, and vulgar (koccai) Tamil. The basis for these distinctions are phonological, morphological, and lexical. Zvelebil made in-depth studies of the Erode (Western), Tuticorin (Southeastern) and Ramnad (Southern) dialects, as well as more cursory studies of the other dialects mentioned, summarizing the salient features.

Zvelebil's scheme is more or less in agreement with Andronov's monograph on Tamil dialects (1962), an important study and perhaps the only comprehensive one of Tamil dialects in existence. Andronov begins by reviewing references to dialect differences in Tolkaappiyam, Nannuul, and other ancient sources, and continues by differentiating between social dialects (Br, NBr, and Harijan) and geographical dialects. Data concerning Tamil dialects are contrasted with those from literary Tamil. He gives characteristic differences by which various geographical dialects are clearly marked: retroflex /r/ replaced by /y/ in the "northern" dialect; present tense morpheme in /-t/ in the "southern" dialect; retroflex /r/ replaced by retroflex /l/ in the "western" dialect, etc. The features of Ceylon dialect are contrasted with literary Tamil. In discussing social dialects, Andronov examines Brahman dialects (especially those of urban intelligentsia), and contrasts them with "middle" and "lower" caste dialect forms, and notes substantial uniformity among all Brahmin dialects. Data is often taken from pronominal forms, both with regard to social and geographical dialects, since these forms seem to display significant amounts of variation. Andronov's theory of dialectology rests on differentiating phonological, morphological and lexical idiosyncracies. He notes the uniformity of the Brahman dialects as contrasted with the heterogeneity of the nonBrahman dialects. An important bundle of isoglosses between Ceylon dialects and mainland dialects is noted by both Zvelebil and Andronov.[18] Zvelebil notes, in conclusion to his article on two dialects of Ceylon, that there are two subdialects of a single dialect, Ceylonese Colloquial Tamil. The evidence is primarily phonological, but supported by morphological and lexical evidence.

Dialects of Telugu. Telugu dialect work has been scanty compared to work on the other major Dravidian languages. Kelley (1969) gives an overview of the Telugu dialect situation. The Telangana dialect (interior districts) is heavily influenced by Urdu lexically, and speakers regard their dialect as lacking in prestige, as do the speakers in the southwest.

Bh. Krishnamurti (1962) has done lexical studies of agriculture and handwork terminology in Telugu. In his work some bundles of isoglosses emerge, separating the two northern coastal districts of Srikakulam and Vishakapatnam and part of E. Godaveri (old Kalinga Kingdom). A second bundle sets off Rayalseema, Nellore and adjacent parts of Guntur, corresponding to some physical and old political boundaries; other dialects reflect fluctuating political boundaries of the past with some unclear and transitional areas. Kelley notes a difference between literary and colloquial Telugu, although in recent years colloquial has been displacing literary norms.

Dialects of Kannada. Grierson does not discuss how Kannada differs from the other languages, and there seems to be widespread agreement that Kannada is an autonomous Dravidian language. The number of dialects of Kannada (Kanarese) is, according to Grierson, "comparatively small," the most important dialect being Badaga, spoken in the Nilgiris. Another Nilgiris dialect is Kurumba, which is also spoken in Chanda. Differences between Kurumba and Kannada are reported to be slight, and other dialectal differences unimportant. However, a dialect known as Bijapur, also perhaps spoken by the Golars of the Central Provinces, differs in pronunciation from Kannada. For instance, /a/ is often found where Standard Kannada has final /e/; initial /e/ and /ee/ are replaced by /ya/ and /yaa/, etc. Some nasalization of vowels is also found in Bijapur. Some other morphological differences are also noted.

More recently there have been studies by both westerners and Indian scholars on the dialects of Kannada. The best overall statements are to be found in the work of D. N. Shankara Bhat (1967-8) who shows a three-fold caste difference and three-fold geographic differences in the dialects of Mysore district. He lists processes shared or not shared by various dialects and Standard Kannada (e.g. retroflex cluster reduction, palatalization, metaphony, etc.)

Zvelebil (1970) states that besides the dichotomy in Kannada between literary/educated speech and colloquial speech, the latter having at least three social dialects (Brahmin, non-Brahmin and Harijan), there are three major regional dialects of Kannada: Dharwar, Bangalore, and Mangalore, which are equivalent to what others have considered North Canara, South Canara, and Old Mysore State.

Dialects of Malayalam. Malayalam is generally agreed to have been a dialect of Tamil until sometime between the tenth and the thirteenth centuries, at the end of which time a written form of the language emerged which was definitely different from Tamil. Colloquial and literary Malayalam differ enough to deserve mention from Grierson. As for dialects, a form called Yerava spoken in Coorg is reported by Grierson, but it is noted that "we have no information about the existence of definite Malayalam dialects." (*LSI*, vol. IV:348) However, it is evident even from Grierson's treatment of Malayalam that some dialects must exist, e.g. his discussion of the remnants of the personal terminations of verbs found supposedly in the Malayalam used in the Laccadives, and among the Moplahs of South Canara. No examples of any dialects of Malayalam are given in *LSI*.

A summary of dialect work on Malayalam is given by V. I. Subramoniam (1969). There are, however, few systematic studies yet in existence of the overall Malayalam dialect picture. We only have a scattered series of studies of various dialects. For instance, there is the thesis of M. V. Sreedhar (1964) giving phonological and morphological statements about that dialect. A phonemic sketch of the Nayar dialect, a short article on the South Kerala dialect, and an exploratory note on the Kayavar dialect also exist.[19] V. I. Subramonium has written a phonemic outline of the Palaya dialect of Malayalam.[20] Popular treatments on the heterogeneous nature of Malayalam have been written by various people. Since the article by Subramonium was written, Zvelebil outlines three distinct territorial dialects: South Kerala, Central Kerala, and North Kerala. Apart from the regional dialects, there are also caste and communal dialects (Namboodiri, Nayar, Moplah, Pulaya, Nasrāni, etc.). Paul Friedrich seems to think there is a four-way division in Malayalam among the speech of Brahmins, Nayars, other touchable castes, and untouchables.[21]

Dialect studies of other Dravidian languages. We have already discussed the status of some speech forms thought by some to be dialects of some languages and by others to be independent languages. We have already noted whom we agree with in various cases. Some of the non-literary languages seem to have dialects worth noting.

Tulu. Tulu has at least a dichotomy between Brahmin and non-Brahmin dialects (L. V. Ramaswami Aiyar, 1932) and some other dialectal differences may also exist. Koraga (Shankara Bhat, 1968a) is now thought to be an independent language or languages.

Gondi. Gondi, as mentioned earlier (Tyler, 1969) seems to be a continuum of mutually intelligible dialects, decreasing in mutual intelligibility with distance. As Zvelebil notes, "The problems of the relationship among different Gondi dialects, the questions of setting up isoglosses, of positing main dialect groups and sub-groups, and finally the reconstruction of Proto-Gondi are so complicated and far-reaching, that obviously Gondi linguistics as such will become in the future a most important and fascinating part of Dravidology."[22]

3.3. The classification of the Munda languages

3.3.0. Introduction

The Munda (older term Kol) languages are spoken principally by tribal groups in Central and Eastern India. It is known that these languages have been in India since before the arrival of the Aryans, and almost certainly once occupied a larger territory in India than they do now. According to Norman Zide, information about these languages has existed in the West since the early 19th century.[23] Owing to the tribal nature of the groups speaking many of these languages, it was originally difficult to correctly identify the language family of some of them, and they have on occasion been confused with tribal Dravidian and Austro-Asiatic non-Munda languages. Comprehensive data on many of the Munda languages is even today lacking, rendering genealogical studies of the language family difficult. This is further complicated by the relative isolation of many of the groups speaking these languages.

It is not the place here to survey the full history of the classification of the Munda languages. We do examine three sources, the *LSI*, Pinnow's monograph on Kharia (Pinnow, 1959), and the *CTL5* article on Munda by Norman Zide. Those interested in the full details of past studies on Munda should consult the bibliographies in those works, as well as in Stampe's bibliographic studies (Stampe, 1965).

3.3.1. The LSI on Munda

According to Zide, "the *LSI* gathered information on the Munda languages and assembled a bibliography of earlier materials. The views and terminology of the *LSI* still constitute most of the common knowledge on Munda, particularly in India." (Zide, 1969:412)

Grierson, after Peter W. Schmidt, considers the Munda languages to constitute a portion of the Austro-Asiatic division of the "Austric Family", the other division of this family being the "Austro-Nesian languages" consisting of the languages of Madagascar, Indonesia, and the islands of the Pacific. The "Austro-Asiatic" languages were stated to be distributed over "Nearer and Further India." (*LSI*, vol. 1, p. 32). The Austro-Asiatic branch is in turn divided by Grierson into a Mōn-Khmēr branch spoken in Assam, Burma, and parts of Indo-China, the main languages of which are Mōn, Khmēr, Palareng, Wa, Khāsī, and Nicobarese. Of these, only Khāsī, spoken in Assam, was seriously examined in the *LSI*. Nicobarese was considered by Grierson to constitute a connecting link between the Munda languages and Mōn (*LSI*, vol. 1, p. 33).

Grierson notes the existence of eight autonomous Munda languages, but considers several of these to have numerous dialects. He uses the cover term Khērwārī to designate a number of dialects spoken at the northeastern end of the Central Indian plateau as well as in adjacent areas. The most important of these are Santālī, Muṇḍārī, Hō, Bhumij and Korwā. Other major Munda languages which he notes are Kūrkū, Khāriā, Juāng, Savara, and Gadabā. Kūrkū is considered to have two dialects, Muwāsī and Nahālī (*LSI*, vol. 1, p. 34).

3.3.2. Pinnow on Munda

Heinz-Jürgen Pinnow, in the introduction to his important 1959 monograph, *Versuch einer historischen Lautlehre der Kharia Sprache*, provides a classification of the entire Austro-Asiatic family of languages. This family is divided into two divisions, a West-Obergruppe having Munda and Nihali as its two components, and an East-Obergruppe encompassing all of the remaining Austro-Asiatic languages. The Munda (or Northwest) component of the "West-Obergruppe" of Austro-Asiatic is asserted to have East, West, Central, and Southern subgroups. The east (or Kherwari) subgroup includes Santali, Muṇḍari, Ho, Bhumij, Birhoṛ, Koḍa, Turi, Asuri and Korwa; the west subgroup consists essentially of Kurku; the central subgroup contains Khaṛia and Juang; while the southern group consists of Sora, Pareng, Gutob and Remo. Pinnow has subsequently discussed the relationship of Munda to the other Austro-Asiatic languages in other articles.[24]

3.3.3. Zide on Munda

Probably the most authoritative source of information concerning the internal relations among the Munda languages is to be found in Zide, 1969. Zide's classification is based on extensive historical reconstruction of the Munda family. Zide's reconstruction of the family is shown in the Stammbaum given in Figure 42.

3.4. The classification of the Tibeto-Burman languages

3.4.0. Introduction

The Tibeto-Burman family of languages, considered by many to be a branch of a larger Sino-Tibetan family, is, in terms of number of languages, the largest of any spoken in South Asia. These languages cover a vast territory, ranging from Jammu and Kashmir in the west to Assam, Indo-China, and parts of China in the east. The groups speaking many of these languages are highly isolated, and only preliminary data exist for many of them. The total number of these languages is large--some 300 are reported by Shafer--and the comparative analysis of data from even a fraction of them is extraordinarily difficult. The inaccessibility of many of the groups speaking these languages makes it difficult to accept the comprehensiveness of the presently accepted inventory of these languages. It is quite possible that as yet unrecorded Tibeto-Burman languages will be discovered in the future.

The lack of much available data for many of these languages coupled with our own lack of expertise concerning them make it impossible to fully discuss the history of classificatory studies of them. The field

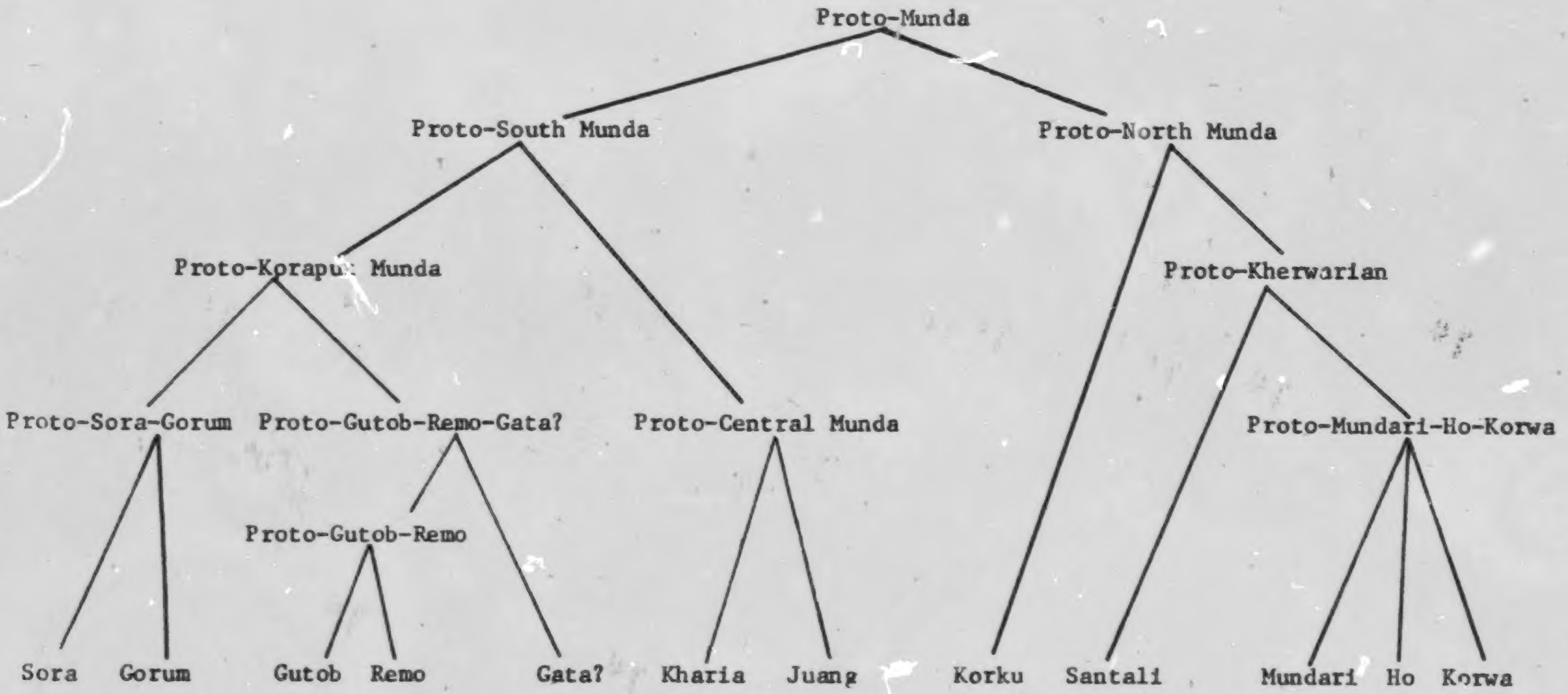

Proto-Munda
Proto-South Munda
Proto-North Munda
Proto-Koraput Munda
Proto-Kherwarian
Proto-Sora-Gorum
Proto-Gutob-Remo-Gata?
Proto-Central Munda
Proto-Mundari-Ho-Korwa
Proto-Gutob-Remo
Sora
Gorum
Gutob
Remo
Gata?
Kharia
Juang
Korku
Santali
Mundari
Ho
Korwa

Figure 42. The Munda languages (after Zide 1969:412).

has fostered numerous controversies concerning the relations among these languages which we are unable to resolve here. We merely provide here, for informational purposes, an overview of the classification of these languages as provided by Shafer, 1955, and as taken up by Miller in his survey article in *CTL5*.

3.4.1.0. Shafer on the "Tibeto-Burman" languages

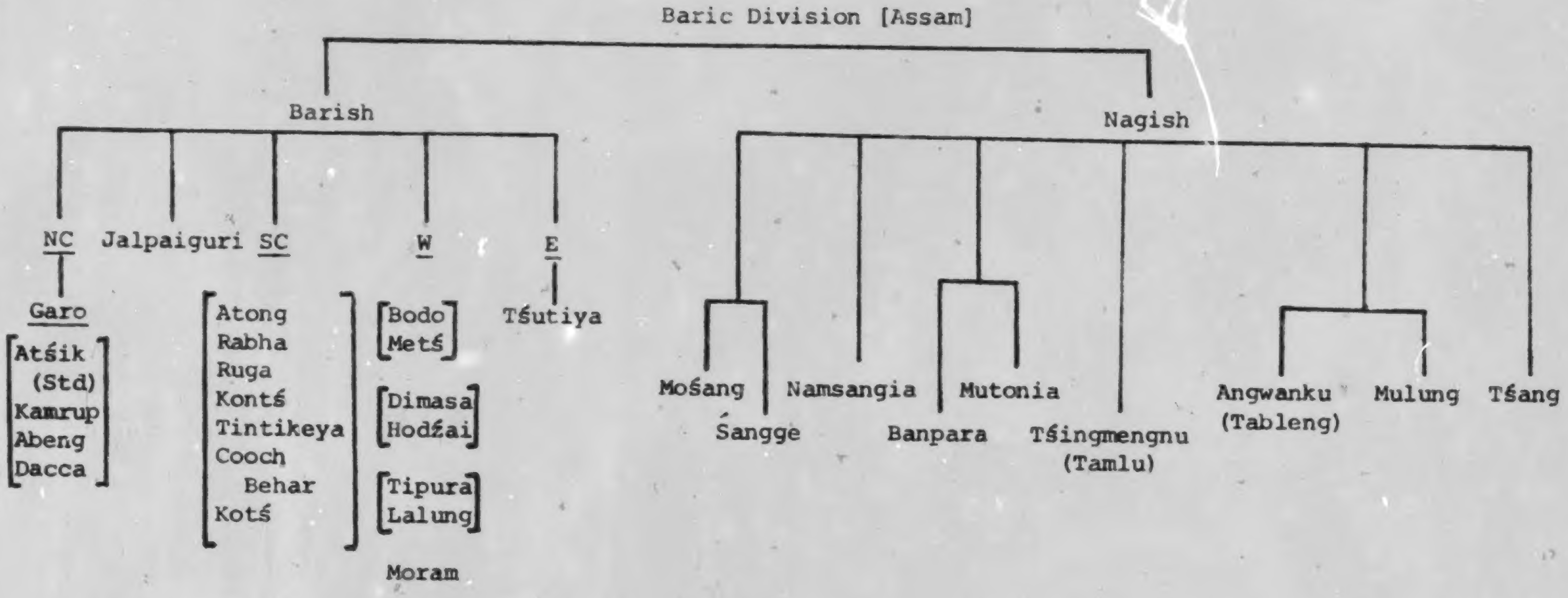
Those groups of languages commonly referred to as the "Tibeto-Burman" languages, a usage well established as far back as the *LSI*, do not constitute an autonomous group of languages in Shafer, 1955. Rather these languages constitute two of a total of six primary divisions of the Sino-Tibetan family. These divisions are (1) the Karenic [Central and Southern Burma], (2) the Baric [Assam], (3) the Burmic [Indo-Burman frontier, Burma, Indo-China, E. Tibet, S. W. China], (4) the Bodic [stretching from the Western Himalayas through Nepal and into Assam, Tibet and Western China], (5) the Daic [West China, Tonkin, Laos, Thailand, parts of Burma], and (6) the Sinitic [China]. The last two of these have been grouped together by some scholars as Sino-Thai family of languages. Those languages commonly thought of as "Tibetan" are included under number four above. It is to be noted that there is neither an autonomous "Tibeto-Burman" nor a "Sino-Thai" family in Shafer, 1955, but rather a single Sino-Tibetan family branching off into the six above-mentioned "divisions".[25] Among languages spoken in South Asia there are representatives of three of these divisions, the Baric, the Burmic, and the Bodic.

3.4.1.1. The Baric Division

According to Shafer, the Baric Division consists of two "sections," the "Barish" and the Nagish. The first of these is in turn, broken down into five "sections," North Central, Jalpaiguri, South Central, West, and East. The other of these, the Nagish, is divided into six sections, with four of those containing two languages each. The structure of the Baric Division is shown below in Figure 43.

3.4.1.2. The Burmic Division

Shafer divides the Burmic languages into eight sections (Figure 44). Of these, only one, the Kukish, is significantly represented in South Asia. This section is an extraordinarily complex one, having a very large number of languages, and being sufficiently unknown as to make its analysis difficult. Shafer divides this section into 12 autonomous


Baric Division [Assam]
Barish
Nagish
NC
Jalpaiguri
SC
W
E
Garo
Atśik
(Std)
Kamrup
Abeng
Dacca
Atong
Rabha
Ruga
Kontś
Tintikeya
Cooch
Behar
Kotś
Bodo
Metś
Dimasa
Hodźai
Tipura
Lalung
Moram
Tśutiya
Mośang
Śangge
Namsangia
Banpara
Mutonia
Tśingmengnu
(Tamlu)
Angwanku
(Tableng)
Mulung
Tśang

Std = Standard
NC = North Central
SC = South Central
W = West
E = East

Figure 43. The Baric Division (After Shafer 1955)

branches (Figure 45), of which the most important in South Asia is the Northern Naga.

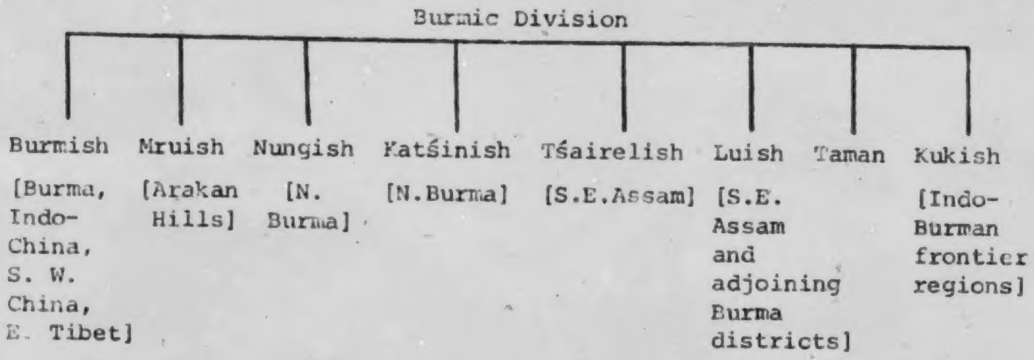

Burmic Division
Burmish
[Burma, Indo-China, S. W. China, E. Tibet]
Mruish
[Arakan Hills]
Nungish
[N. Burma]
Katšinish
[N.Burma]
Tšairelish
[S.E.Assam]
Luish
[S.E. Assam and adjoining Burma districts]
Taman
Kukish
[Indo-Burman frontier regions]

Figure 44. The Burmic Division (after Shafer, 1955)

3.4.1.3. *The Bodic Division*

The Bodic Division of the Sino-Tibetan language family is of complexity equal to that of the Burmic. Shafer posits 11 sections (Figure 46) of this Division, the Bodish, West Himalayish, West Central Himalayish, East Himalayish, Newarish, Digarish, Midźuish, Hrusish, Dhimalish, Miśingish and Dzorgaish respectively. Several of these sections are themselves of great complexity. The Bodish section (Figure 47) of Bodic contains a large number of Himalayan languages, including the literary and standardized varieties of Tibetan. The West Himalayish section (Figure 48) is also quite complex, and is divided by Shafer into five branches, North-Northwest, Northwest, Almora, Dźanggali, and Eastern. The East Himalayish section (Figure 49) of Bodic is likewise divided into two major branches, Eastern and Western, both of which themselves have subdivisions.

3.4.2. *Miller on the Tibeto-Burman languages*

Miller in his 1969 article in *CTL5* attempts to summarize the results of classificatory studies of the "Tibeto-Burman" prior to then, as well as to note the major problems and critical needs in this field. Miller, 1969 accepts Shafer, 1955 as a basic typological reference point, but departs from it in several ways, incorporating the results of at least a decade of research subsequent to Shafer, 1955. We will here merely note the major deviations of Miller's article from that stated earlier by Shafer.

The most important of these deviations concerns the primary

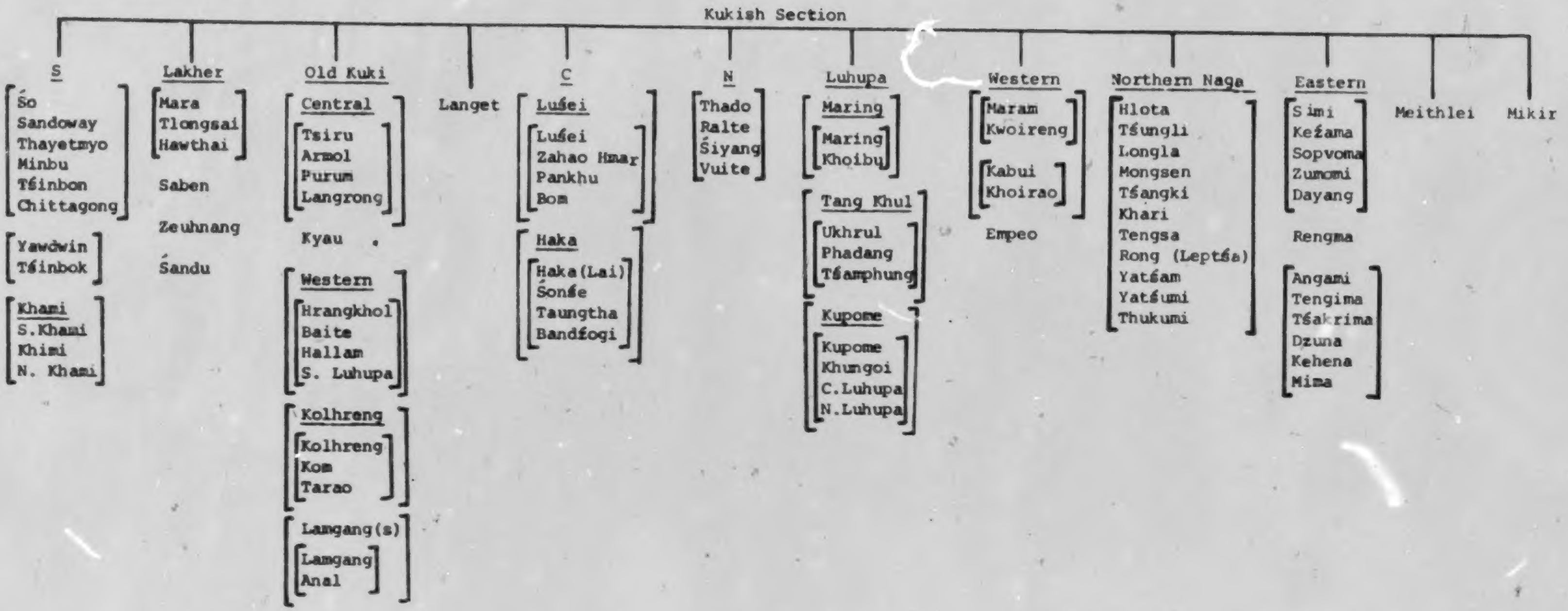

Kukish Section
S
Śo
Sandoway
Thayetmyo
Minbu
Tśinbon
Chittagong
Yawdwin
Tśinbok
Khami
S.Khami
Khimi
N. Khami
Lakher
Mara
Tlongsai
Hawthai
Saben
Zeuhnang
Śandu
Old Kuki
Central
Tsiru
Armol
Purum
Langrong
Kyau
Western
Hrangkhol
Baite
Hallam
S. Luhupa
Kolhreng
Kolhreng
Kom
Tarao
Langang(s)
Langang
Anal
Langet
C
Luśei
Luśei
Zahao Hmar
Pankhu
Bom
Haka
Haka(Lai)
Śonśe
Taungtha
Bandźogi
N
Thado
Ralte
Śiyang
Vuite
Luhupa
Maring
Maring
Khoibu
Tang Khul
Ukhrul
Phadang
Tśamphung
Kupome
Kupome
Khungoi
C.Luhupa
N.Luhupa
Western
Maram
Kwoireng
Kabui
Khoirao
Empeo
Northern Naga
Hlota
Tśungli
Longla
Mongsen
Tśangki
Khari
Tengsa
Rong (Leptśa)
Yatśam
Yatśumi
Thukumi
Eastern
Simi
Keźama
Sopvoma
Zumomi
Dayang
Rengma
Angami
Tengima
Tśakrima
Dzuna
Kehena
Mima
Meithlei
Mikir

Figure 45. The Kukish Section of Burmic
(after Shafer 1955)

segmentation of Tibeto-Burman into "immediate constituents". Shafer had posited a six-way branching of Sino-Tibetan into Daic, Sinitic, Bodic, Baric, Karenic and Burmic, with no special affinities among any combinations of these. Miller, on the other hand, seems to accept the existence of an autonomous Tibeto-Burman, which would presumably be a first-order constituent of Sino-Tibetan. Miller divides this Tibeto-Burman entity down into a Tibetan component and a Burmese component. Miller's "Tibetan" component fairly closely corresponds to Shafer's Bodic Division, except that Shafer's Bodish section appears as Miller's Tibetan section, and Miller has added a tenth section, the Midźuish, to Shafer's nine. Miller's Burmese component (Figure 50) differs substantially from Shafer's Burmic Division. Whereas Miller seems to suggest a binary division of Burmese into a Burmese Section and a Kuki section,[26] Shafer (Figure 44) gives an eight-way division.

3.4.3. Problems in the taxonomy of Tibeto-Burman languages

It has become obvious to us that at the present time the Tibeto-Burman languages pose extraordinarily difficult problems for classificatory analysis. The number of such speech forms is vastly larger than for any other family of languages in South Asia. With few exceptions these languages are very inadequately described in the scholarly literature. The overwhelming majority of them lack literary forms and documentable textual traditions. The areas in which many of these languages are spoken are highly inaccessible. Moreover, those scholars who have investigated these languages differ among themselves in the criteria to be used in their comparative analysis. With so little to go on, the chaotic situation which currently exists concerning the mutual relations and affinities among those languages is hardly surprising.

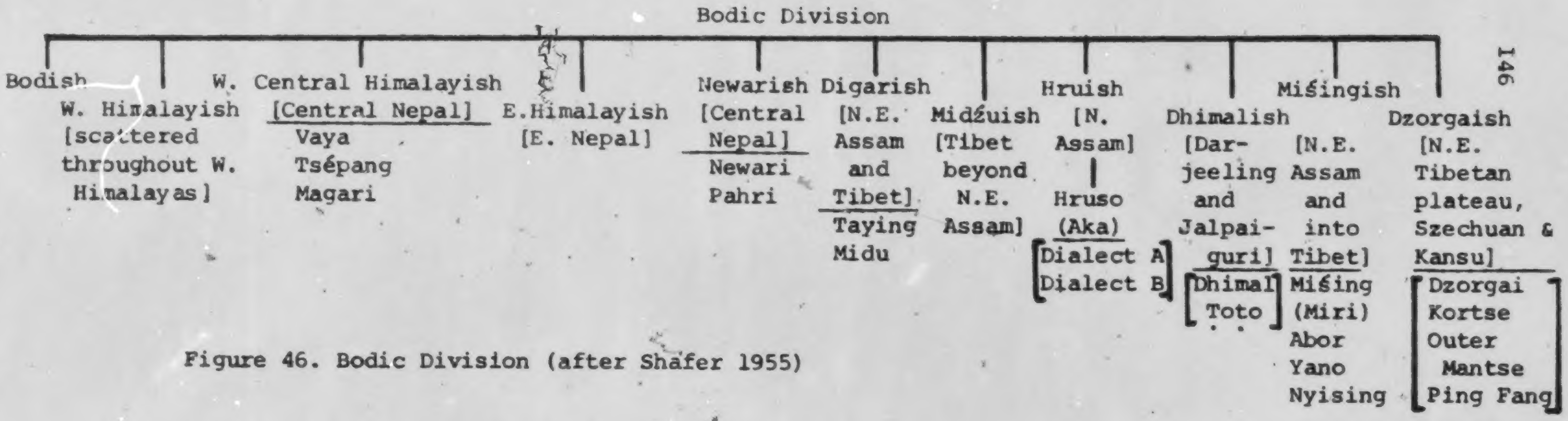

Bodic Division
Bodish
W. Himalayish
[scattered
throughout W.
Himalayas]
W. Central Himalayish
[Central Nepal]
Vaya
Tsépang
Magari
E.Himalayish
[E. Nepal]
Newarish
[Central
Nepal]
Newari
Pahri
Digarish
[N.E.
Assam
and
Tibet]
Taying
Midu
Midźuish
[Tibet
beyond
N.E.
Assam]
Hruish
[N.
Assam]
Hruso
(Aka)
Dialect A
Dialect B
Dhimalish
[Dar-
jeeling
and
Jalpai-
guri]
Dhimal
Toto
Miśingish
[N.E.
Assam
and
into
Tibet]
Miśing
(Miri)
Abor
Yano
Nyising
Dzorgaish
[N.E.
Tibetan
plateau,
Szechuan &
Kansu]
Dzorgai
Kortse
Outer
Mantse
Ping Fang

Figure 46. Bodic Division (after Shafer 1955)

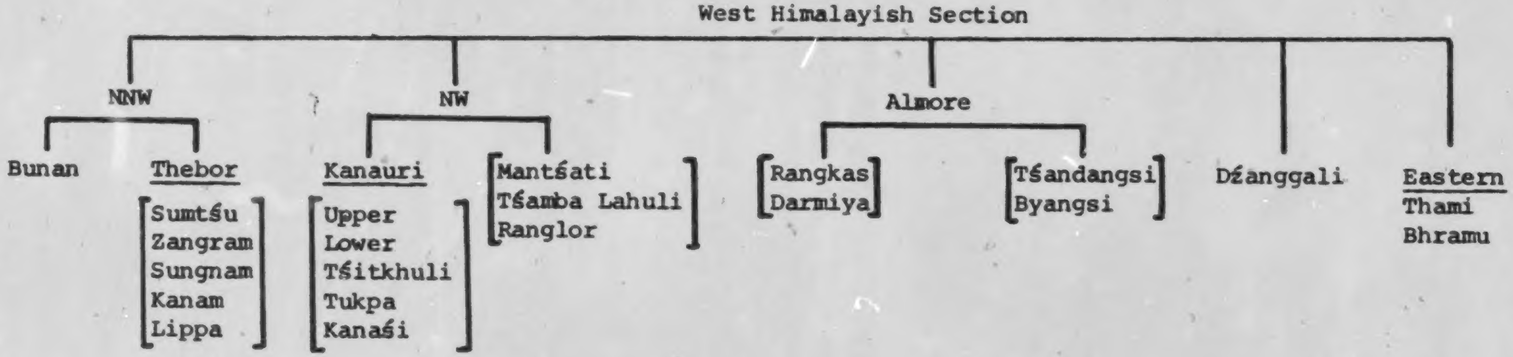

West Himalayish Section
NNW
Bunan
Thebor
Sumtśu
Zangram
Sungnam
Kanam
Lippa
NW
Kanauri
Upper
Lower
Tśitkhuli
Tukpa
Kanaśi
Mantśati
Tśamba Lahuli
Ranglor
Almore
Rangkas
Darmiya
Tśandangsi
Byangsi
Dźanggali
Eastern
Thami
Bhramu

Figure 48. West Himalayish Section of Bodic (after Shafer 1955)

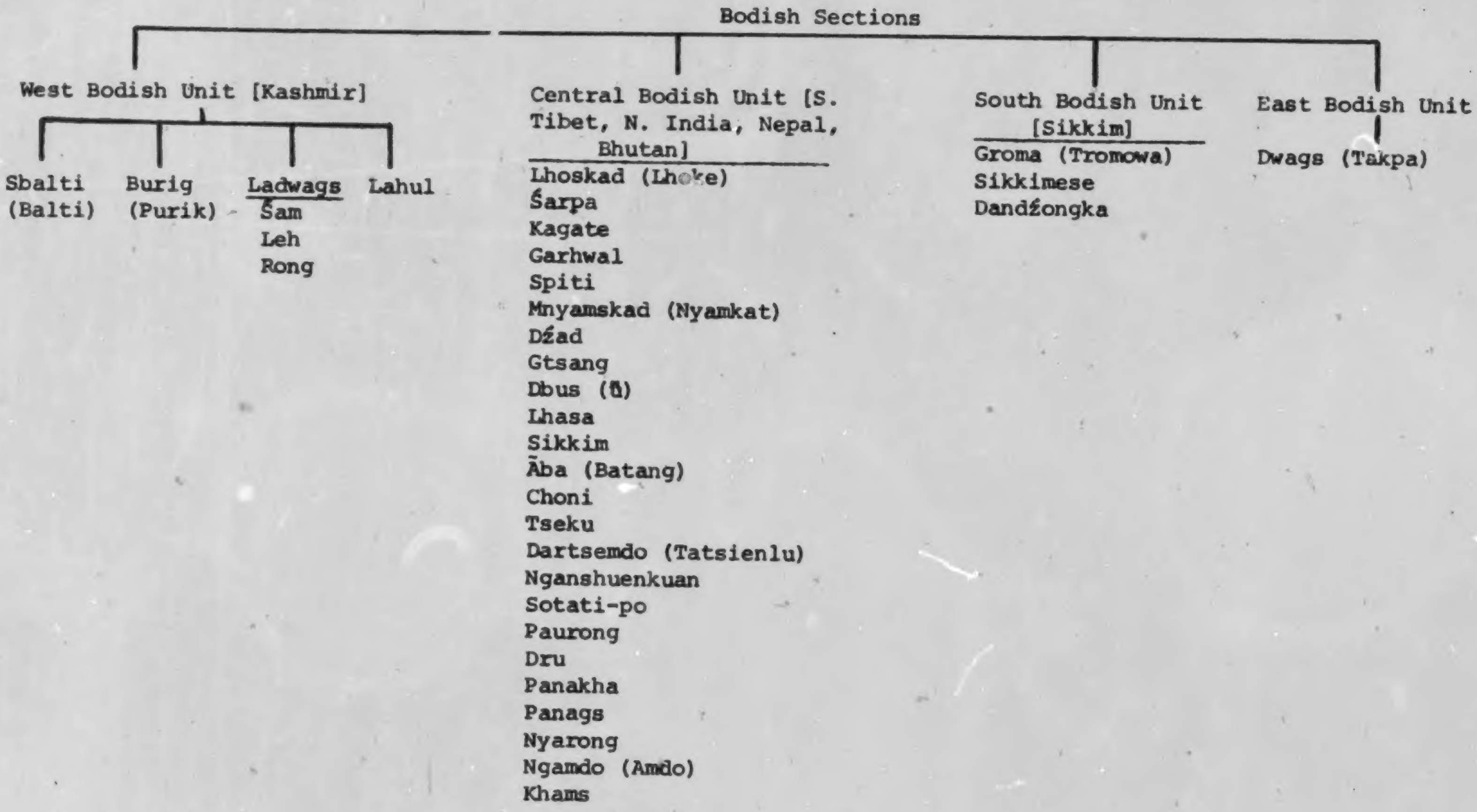

Bodish Sections
West Bodish Unit [Kashmir]
Sbalti (Balti)
Burig (Purik)
Ladwags
Śam
Leh
Rong
Lahul
Central Bodish Unit [S. Tibet, N. India, Nepal, Bhutan]
Lhoskad (Lhoke)
Śarpa
Kagate
Garhwal
Spiti
Mnyamskad (Nyamkat)
Dźad
Gtsang
Dbus (Ü)
Lhasa
Sikkim
Āba (Batang)
Choni
Tseku
Dartsemdo (Tatsienlu)
Nganshuenkuan
Sotati-po
Paurong
Dru
Panakha
Panags
Nyarong
Ngamdo (Amdo)
Khams
South Bodish Unit [Sikkim]
Groma (Tromowa)
Sikkimese
Dandźongka
East Bodish Unit
Dwags (Takpa)

Figure 47. The Bodish Section of Bodic (after Shafer 1955)

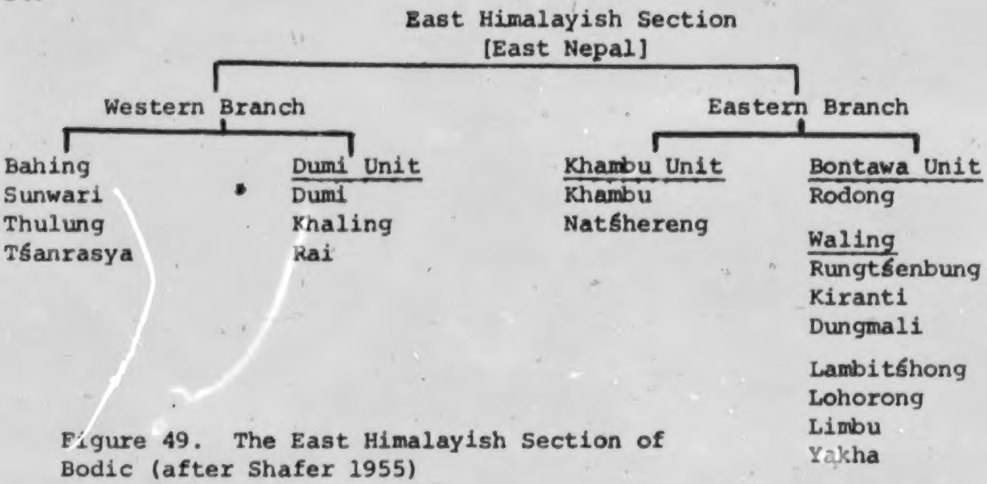

East Himalayish Section
[East Nepal]
Western Branch
Eastern Branch
Bahing
Sunwari
Thulung
Tśanrasya
Dumi Unit
Dumi
Khaling
Rai
Khambu Unit
Khambu
Natśhereng
Bontawa Unit
Rodong
Waling
Rungtśenbung
Kiranti
Dungmali
Lambitśhong
Lohorong
Limbu
Yakha

Figure 49. The East Himalayish Section of Bodic (after Shafer 1955)

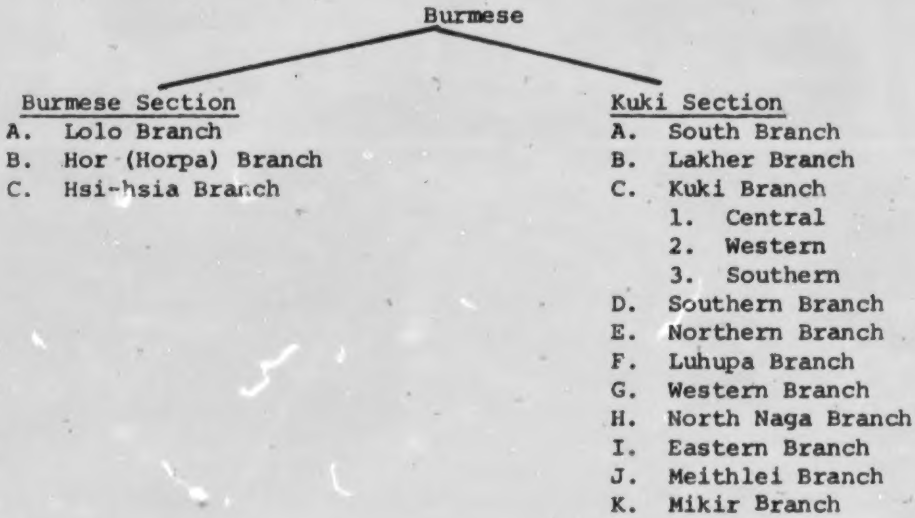

Burmese
Burmese Section
A. Lolo Branch
B. Hor (Horpa) Branch
C. Hsi-hsia Branch
Kuki Section
A. South Branch
B. Lakher Branch
C. Kuki Branch
1. Central
2. Western
3. Southern
D. Southern Branch
E. Northern Branch
F. Luhupa Branch
G. Western Branch
H. North Naga Branch
I. Eastern Branch
J. Meithlei Branch
K. Mikir Branch

Figure 50. The "Burmese" languages (after Miller 1969)

NOTES: CHAPTER 3

1. An excellent summary of early attempts to describe the Indo-Aryan languages of South Asia is to be found in the *Census of India 1961*, Volume I, Part XI-C(i), *Inquiries into the Spoken Languages of India from Early Times to Census of India 1901* (Language Division, Office of the Registrar General, India). This work cites works as early as H. T. Colebrooke's "On the Sansirit and Pracrit Languages" (*Asiatick Researches*, vol. VII, Art. VII [1801], pp. 199ff) and reports by William Carey as pointing out the existence of a North Indian group of related vernacular languages. Grierson in *LSI*, vol. I, pp. 1-17 provides an excellent summary of pre-19th century accounts of the linguistic situation in India.

2. This is certainly true of many of the so-called dialects of Hindi, as, for example, Bhojpuri, Awadhi, Rajasthani, Maithili, etc. Each of these has a considerably old literature which could have served as a focal point in the emergence of a modern standardized language.

3. Cf. Beames, 1960:48-54.

4. Grierson's conclusions concerning the historical development of Indo-Aryan are provided in *LSI*, vol. 1, pp. 115-33.

5. Grierson is, however, extremely perceptive in noting the difficulties inherent in trying to differentiate between languages and dialects.

> In the course of the Survey, it has sometimes been difficult to decide where a given form of speech is to be looked upon as an independent language, or as a dialect of some other definite form of speech. In practice it has been found that it is sometimes impossible to decide the question in a manner which will gain universal acceptance. The two words 'language' and 'dialect' are, in this respect, like 'mountain' and 'hill'. . . 'language' and 'dialect' are often used in the same loose way. In common use we may say that, as a general rule, different dialects of the same language are sufficiently alike to be reasonably well understood by all whose native tongue is that language, while different languages are so unlike that special study is needed to enable one to understand a language that is

> not his own. This is the explanation of the Century Dictionary, but the writer adds that 'this is not the essential difference,' and no where is this proviso needed more than in considering the Aryan languages of Northern India. There, mutual intelligibility cannot always be the dividing factor, for the consideration is obscured by the fact that between Bengal and the Panjab every individual who has received the slightest education is bilingual. In his own home, and in his own immediate surroundings he speaks a local idiom, but in his intercourse with strangers he employs or understands some form of that great lingua franca,--Hindī or Hindūstānī . . . . The differentiation of a language does not necessarily depend on non-intercommunicability with another form of speech. There are also other powerful factors to be considered if we are to look at the subject from a scientific point of view. First and foremost, there is. . .grammatical structure. . . . There is [also] another factor which exercises influence in this differentiation. It is nationality. (*LSI*, vol. 1, pp. 23-4)

6. Cf. Brass, 1974:51-116.

7. Cf. Dimock, 1960.

8. For further information on the classification of the Dardic languages see *LSI*, vol. 1, pp. 108-14, G. A. Grierson, *The Piśāca Languages of North-western India* (London: The Royal Asiatic Society, 1906 [reprinted Delhi: Munshiram Manoharlal, 1969]), and Braj B. Kachru, "Kashmiri and the Other Dardic Languages," in *CTL5*, pp. 284-306.

9. Cf. Caldwell, 1856:3-6.

10. "The weight of comparative evidence. . .is in favor of considering Telugu as an off-shoot of the Central Dravidian branch of proto-Dravidian. . . . Since it has also several exclusive features in common with South Dravidian in phonology rather than in morphology, it may be considered that Telugu has been in intimate geographical contact with the members of South Dravidian from a very remote past. The morphological evidence puts it rather conclusively with Central Dravidian." (Krishnamurti 1961:269)

11. Cf. Andronov, 1963, 1970:23.

12. Andronov, 1970:23.

13. Cf. Diffloth, 1968:13. The originial material is contained in F. H. Gravely, *Gramophone Records of the Languages and Dialects of the Madras Presidency*, Texts of the passages. Government Press, Madras, 1927.

14. Diffloth, 1968:14.

15. Diffloth, 1968:14fn.

16. Shankara Bhat, 1968.

17. Shankara Bhat, 1968:291.

18. Zvelebil, 1966 and Andronov, 1962.

19. K. Kuññuṇṇi Raajaa, "Nasal Phonemes of Malayalam," *IL*, vol. 21 (1960), 90-96; C. R. Sankaran and A. C. Sekhar, "The dialect of the extreme South of Kerala," *Bulletin of the Deccan College Research Institute*, vol. 7 (1946), pp. 220f.; and A. C. Sekhar, "A Note on the Kayavar Dialect," *Bulletin of the Deccan College Research Institute*, vol. 10 (1950), pp. 47f.

20. V. I. Subramoniam. "Phonemic Outline of a Dialect of Malayalam," *IL*, vol. 23 (1962), pp. 99-116.

21. Cf. Friedrich, 1961.

22. Zvelebil, 1970:17.

23. For a detailed bibliography of early research on Munda languages see Pinnow, 1959;459-89 as well as *Inquiries in the Spoken Languages of India from Early Times to Census of India 1901* [*Census of India 1901*, vol. 1, part XI-C(i), Language Monographs], New Delhi, 101-6.

24. See particularly Pinnow, 1963.

25. Shafer (1955:94-6) discusses what he considers to be the unnatural division of the Sino-Tibetan family of languages into a "Chinese-Siamese" sub-family and a "Tibeto-Burman" one. He considers this division to be the accidental result of the different traditions out of which "Chinese-Siamese" and "Tibeto-Burman" linguistic studies respectively developed. He notes that "Henri Maspero, the last great scholar to hold the "Siamese-Chinese" division, was a product of the École Francaise d'Extrême-Orient at Hanoi. . . [and] knew practically nothing of the "Tibeto-Burman" languages which had been studied for some time primarily under the patronage of the British in India, and so Maspero naturally accepted "Tibeto-Burman" as a sub-family on the authority of those who had been studying those languages." Similarly Sten Konow, the Norwegian scholar engaged by the British Government to handle non-Indo-Aryan languages for the *LSI*, knew virtually nothing of the "Siamese-Chinese" languages, and did nothing to challenge the legitimacy of that designation.

26. Miller (personal communication) has indicated that the omission of any discussion of Shafer's Mruish, Nungish, Katśinish, Tśairelish, Luish, and Taman sections in his 1969 article was through oversight and does not represent a judgement as to the internal structure of the Burmish Division.

Chapter 4

South Asia as a Linguistic Area

4.0. Introduction

In the previous two chapters we attempted to outline the major languages and language families of South Asia, as well as to discuss the criteria which were used in determining their limits. It has become obvious that there are no absolute standards by which language classification can be carried out, and that all taxonomies of language varieties involve a subjective emphasis on some aspects of linguistic structure over others. We have seen that genetic classifications of sets of speech varieties vary according to how one defines terms such as "language" and "dialect."

In the case of South Asia, the vast majority of classificatory descriptions of language varieties attempt to express generalizations via some form of Stammbaum approach, and the consequences of adopting this approach are many. Languages have been grouped together as members of a single family because it is possible to postulate historical rules of sound change relating their forms to earlier ancestral forms, a result being that the "core" of these languages is inevitably defined as that portion of them which admits such reconstruction. All else must be considered peripheral for taxonomic purposes.

Loan words are a case in point. Because they frequently do not observe the same phonological "laws" as other forms in a language, they are frequently excluded from the core of items whose comparison serves as the basis for genetic groupings. Reconstructions and genetic classifications are based on the comparison of "inner cores" of language varieties which remain after stripping away linguistic features likely to have resulted from the influence of other speech varieties. This procedure leads to an arbitrary bias in the derived classifications. Languages in close proximity invariably influence each other in some way: lexical and phonological borrowings frequently occur; stress patterns may be altered; entire syntactic constructions may be adopted; or morphological distinctions may be added, lost, or changed. Yet these phenomena, resulting from the proximity of language varieties, are precisely those which are characteristically removed from the "cores" of speech forms being compared.

In order to explain the relations holding between language varieties which result from their proximity rather than from their common parentage, other methods need to be used. It has been commonly observed that language varieties in close proximity exhibit shared linguistic features which are not likely to have developed independently in the separate languages. Cases of such convergence are to be found in the Balkans, the Caucasus, areas of Eastern Europe, parts of North America (vis-à-vis American Indian languages), and South Asia.[1] Such an area has been called a "linguistic area" (German *Sprachbund*) and is usefully defined by Emeneau as "an area which includes languages belonging to more than one family but showing traits in common which are found not to belong to other members of [at least] one of the families" (Emeneau, 1956:16fn.).[2] Examples of linguistic convergence in South Asia have been known for hundreds of years,[3] although the postulation of a full-fledged linguistic area is of fairly recent origin.

Linguistic areas are of interest from a number of points of view. Synchronically, they provide us with an alternative to the genetic Stammbaum model for classifying language varieties. They also raise questions about the historical contact between diverse linguistic groups as well as about the direction of borrowing of linguistic items between codes in contact. This in turn raises questions concerning the mechanisms by which linguistic convergence takes place and the social conditions which cause these linguistic changes to occur. All of this, of course, has implications for any reconstruction of the prehistory of South Asia.

4.1. Linguistic Bases of the South Asian Linguistic Area

There are a number of linguistic features which have been cited in defining a South Asian linguistic area. The literature in which these features are discussed often focuses on the historical processes which lead to their dissemination over a wide range of languages. Many of the features identified as being areal have been observed in grammars dating back at least as far as Caldwell's *Comparative Grammar of the Dravidian or South Asian Family of Languages* (Caldwell, 1856). Almost all of these data are drawn from Indo-Aryan and Dravidian, with less taken from Munda and hardly any from Tibeto-Burman. Emeneau first postulated a South Asian linguistic area in 1956, but many of the criteria which he cites in setting up the area were pointed out earlier.

Jules Bloch (Bloch, 1934), although in general seeking to minimize the structural importance of non-Indo-Aryanisms in Indo-Aryan, nevertheless cites a number of significant areal features. Many of these features broadly pertain to a wide variety of languages, while others relate to restricted aspects of a small set of genetically unrelated languages:

1. The existence in Sanskrit of names of tribes of ancient peoples which form rhyming pairs by a process of reduplication plus initial consonant change (e.g., Pulina-Kulina, Kosala-Tosala, Kalinga-Tralinga). This process seems to be Austro-asiatic in origin.

2. The existence of Dravidian loan words in Sanskrit, dating as far back as Ṛgveda, e.g., RV *ulū́khala-* 'mortar', AV *músala-* 'pestle'.

3. The sharing by Santali, a north Munda language, of a number of vocabulary items with certain dialects of Hindi, as well as with Oriya and Bengali. In addition there are several lexical items in Sanskrit which are likely of Munda origin: *tāmbūla* 'betel', *kadala-* 'banana', *bāṇa* 'bamboo arrow'.

4. The adoption of numerous lexical items from Indo-Aryan (particularly Sanskrit) into all Dravidian languages.

5. The presence in Indo-Aryan, Dravidian, and Munda (except Sora) of a series of retroflex consonants which are contrasted with dentals (a contrast not shared by Indo-Aryan's other Indo-European relatives).

6. The development in Sanskrit's phonological system of short *e* and *o*, in addition to long forms of these vowels. This is in accordance with Dravidian and Munda patterns, both of which have short and long *e* and *o*.

Bloch also points out the following areas of morphological similarity among various Indian languages:

> 7. The consistent use in both Dravidian and Indo-Aryan of suffixes, and deemphasizing of the use of prefixes and infixes.
>
> 8. The absence in both Dravidian and Indo-Aryan of preverbs and prepositions as such.
>
> 9. The absence of a dual number (originally present in Sanskrit, and lost in Middle Indic).
>
> 10. Double nominative stems of nouns, the oblique stem admitting of the force of a genitive and of being followed by words more or less emptied of their proper sense.
>
> 11. Personal pronouns having two stems, that of the nominative and that of the direct and indirect object (e.g., H. *maĩ* 'I'; *mujh see* 'from me'; G. *hũ* 'I'; *m-/mar-* 1st sg. obl.; Ta. *naan* 'I', *en-akku* 'to me.'
>
> 12. The existence in the verb of third persons in the form of nouns, i.e. participles or participial stems (e.g. Ta. *irukkiṟ-avar* "he-who is").
> 1 2 2 1
>
> 13. Varying in gender of participles.
>
> 14. Presence of an absolutive construction. (In this construction two independent clauses are linked together by adding a special non-finite verb form after the verbal stem of the main verb in the first of the clauses. Usually the clauses so conjoined have coreferential subjects, and stand in any of a limited number of logical relations to each other, i.e. temporal subordination, cause and effect, adverbial plus main predication.
>
> 15. Absence of morphologically marked degrees of comparison of adjectives.
>
> 16. The independent adoption in Marathi, Oriya and Sinhalese of the Dravidian relative participle to their syntax, an invariable adjective admitting a subject in the nominative in any construction,

This data needs to be interpreted carefully. Although pointing out many areas in which Indo-Aryan has been influenced by non-Indo-Aryan sources, Bloch considers these areas to be essentially outside of the core of features which defines that language family. He states that "remarkable, and in certain cases conclusive as these concordances may be, the evolution of Indo-Aryan has not resulted in denaturalisation" (Bloch, 1934:328); that is to say, even though there have undoubtedly been areas of mutual influence between Dravidian and Indo-Aryan, these areas have not affected the essential linguistic composition of those language families.

Murray B. Emeneau, in his important 1956 article "India as a Linguistic Area," adds new data to the phenomena described by Bloch and postulates the existence of a distinct Indian linguistic area. Emeneau discusses the historical processes of borrowing which must have been involved in the creation of the area and notes that "the end result of the borrowings is that the languages of the two families, Indo-Aryan and Dravidian, seem in many respects more akin to one another than Indo-Aryan does to the other Indo-European languages" (Emeneau, 1956:16). The bulk of Emeneau's data are taken from Dravidian and Indo-Aryan, although he does cite information from Munda where available and relevant.

After recapitulating areal features scattered throughout earlier literature, the two main sets of new data which Emeneau gives in this article involve what he calls the "echo-word" construction (cf. Bloch's point #1 above) and the use of "classifiers" or "quantifiers." Emeneau defines the "echo-word" construction as one "in which a basic word formulated as CVX is followed by an echo-word in which CV is replaced by a morpheme *gi-* or *u-* or the like (or C is replaced by *m-* or the like), and X echoes the X (or VX echoes the VX) of the basic word. The meaning of the echo-word is 'and the like'; e.g. *puli gili* 'tigers and the like'" (Emeneau, 1956:10). According to Emeneau this construction is found in all Dravidian languages, is widespread in Indo-Aryan, and is attested in Munda in at least Sora.[4] Emeneau concludes that "it is clear already that echo-words are a pan-Indic trait and that Indo-Aryan probably received it from non-Indo-Aryan (for it is not Indo-European)" (Emeneau, 1956:10).[5]

Another areal feature proposed in this Emeneau article is the use of "classifiers" or "quantifiers," described as follows:

> In constructions marked by these [quantifiers or classifiers], when a noun is numerated by means of a numeral or a similar word, the construction contains also one of a smallish class of words or morphemes which we can call by either of these terms. The term 'classifier' indicates that there are as many

> classes of nouns as there are classifiers; the term 'quantifier' indicates that in numeration of nouns there is always specification of the type of unit by which the species indicated by the noun is counted. The units indicated are of various kinds, either measured units of nondiscrete entities (e.g. a quart of liquid, an acre of land) or discrete entities as classed by various criteria (e.g. human vs. animal, animate vs. nonanimate, long and thin vs. flat and thin vs. spherical). Such quantifiers are, to be sure, used in probably all languages; English has *a ton of coal, two acres of land, three head of cattle,* etc. But the languages under discussion at the moment are not those in which only nouns denoting nondiscrete entities and a few others are classified or quantified, but those in which all or nearly all nouns are treated thus. Conspicuous as having such systems are Chinese, Japanese, Korean, Vietnamese, Khmer, Thai, Burmese, and Malay. (Emeneau, 1956:10-11).

According to Emeneau, the construction is of Indo-Aryan origin and has spread to Dravidian and Munda. It has been reported in Bengali, Assamese, Oriya, Maithili, and Emeneau adds citations from Marathi, village Hindi and Nepali. In Dravidian, Emeneau shows forms from Kolami (Wardha dialect), Parji, Kui-Kuvi, Kurukh and Malto. He notes that in Munda, Koṛowa, Santali and Mundari the forms are used, but in Sora they are not. He is able to conclude that there

> is a large area of India, especially eastern and central India, with this feature. . . . the construction (so far as India is concerned) is originally Indo-Aryan. It spread thence to the other languages as a total construction consisting of numeral + classifier, and then was elaborated in some of the languages with native material, the native numerals, native morphemes as additional classifiers, etc. (Emeneau, 1956:14).

Since the publication of Emeneau's "India as a Linguistic Area," the number of proposed areal features has gradually increased. Andronov (Andronov, 1964b) lists additional areal features which are considered to have been borrowed from Dravidian into Indo-Aryan and features which were borrowed in the other direction. Of the former type are the simplification of consonant clusters in Indo-Aryan in accordance with Dravidian phonological patterns; the frequent voicing, spirantization or deletion of intervocalic stops, also in accordance with older Dravidian patterns; the presence of a large number of onomatopoetic terms whose formation accords with older Dravidian patterns; the modification of

Indo-Aryan syntactic patterns to those of the Dravidian languages; and the presence of "chains of participles and compound verbs." Andronov also cites a number of grammatical features of Dravidian which he considers to be of Indo-Aryan origin; including

> the loss of short *e* and *o* in Brahui;
>
> the development of nasal vowels and dipthongs of an Indo-Aryan type in Brahui, Kurukh and several other languages;
>
> the development of aspirate consonants in some of the modern Dravidian languages;
>
> the loss of sentence types in which a synthetic negative form of the verb is used, and the adoption of constructions using a special negative word of the Indo-Aryan type;
>
> the loss of personal nouns, widely used in early Dravidian texts, in many modern Dravidian languages;
>
> the development of adjectives and adverbs in most modern Dravidian languages;
>
> and the development of compound and complex sentences.

Andronov attributes a greater degree of linguistic importance to the notion of linguistic area than does Emeneau, going so far as to suggest that convergence of this sort may actually eradicate genetic boundaries between language families. He sees areal convergence as able to gradually weaken genetic lines of demarcation between language families and ultimately render them irrecoverable and meaningless as typological markers.

> . . .the so-called 'genetic' relationship of languages within one family, a remnant of naturalistic conceptions of language, is not primordial and perpetual. It is historic in its nature: it is formed gradually and gradually can it weaken and disappear. In this sense the development of the typological similarity of the modern Indo-Aryan and Dravidian languages can be regarded as a prerequisite or an initial stage in the formation of a new linguistic family. If the direction of their development does not change in the future, the now observed tendency to develop the formal similarity may gain strength and result in the formation of new relationship ties and of a new language family, which will be neither Indo-European, nor Dravidian. (Andronov, 1964b:13).

Although Andronov's article is useful in providing a summary of South Asian areal features scattered throughout the literature, his theoretical conclusions cited above must be rejected. There is at present no significant evidence demonstrating that multi-lingual convergence leads to the eradication of the distinctiveness of the converging codes. Borrowing and assimilation of grammatical features from one language family to another has taken place and undoubtedly will continue to take place in South Asia; but such borrowing takes place in the history of all language families, and contact situations may provide the impetus for much "naturally motivated" sound change in any case. Such adaptation would remove the distinctiveness between codes only if there were a prior loss of the psychological awareness of the autonomy of codes in the minds of their users. There is no evidence to the effect that this is occurring in South Asia; rather, there is much evidence of the solidification of regional standard languages. It is clear then that rather than seeing the emergence of a new pan-Indian language family, we are observing the continual modification of codes which, in a large number of cases, are considered autonomous by their speakers. With an increase in literacy, and standardization of regional codes, the social circumstances which might have led to the loss of consciousness of the distinction between languages have essentially been lost. As such, the continued fusion of language families, which might have gone on during the early period of Dravidian and Indo-Aryan contact, is unlikely to continue.

A recent article by Emeneau (Emeneau, 1969) expands the discussion of onomatopoetic forms as a pan-Indian areal feature. A large portion of his data comes from Kota,[6] and a discussion of onomatopoetics in Kota forms an excellent basis for examining the feature throughout South Asia. According to Emeneau, onomatopoetics occur in Kota in two basic syntactic constructions:

> (A) as a direct quotation followed by a form of the quotative verb *in-* (second stem *iḏ-*) 'to say so-and-so';
>
> (B) as an expander ('adverb') preceding the verb part of a predication.
>
> A. *avn dop iḏr ki· mu·l viṛti·ko·* 'He fell down with the noise of falling' (*dop iḏr* 'saying *dop*'; 21.145).
>
> B. *avn aṭtltr dopn neta·lk viṛteyṯ*. . .'He having fallen from the attic to the ground with the noise of falling *(dopn)*. . .' (25.113).
>
> A. *cadm kordr kordr vadṯ guc guc iḏi·ko·* 'His voice having gone on becoming reduced gradually, he whispered (said *guc guc*)' (10.88).

B. *a·n gucgucn aṛdge·n* 'I will tell you secretly' (*gucgucn* 'whisperingly'). (Emeneau, 1969:275).

In the examples cited above, the purely onomatopoetic elements are *dop* 'the sound made by a falling object' and *guc guc* 'the sound made while whispering'. These constructions represent the formal means by which such onomatopoetic words are integrated into Kota sentences. The onomatopoetic words are a special class of lexical items having describable phonological properties (although the phonology of onomatopoetic items may differ significantly from the normal phonology of the language), and having a restricted range of meanings.[7]

It has been pointed out in scholarly literature that onomatopoetic forms such as those cited above occur in a wide range of South Asian languages.[8] Although the exact phonological realization of onomatopoetic forms varies with the phonological rules of the particular languages, and in the language-specific morphological or syntactic constructions in which they are used, the existence of such forms seems to be an areal trait. The importance of the Emeneau article extends beyond merely pointing out the existence of a construction in a number of languages. Using data derived from an analysis of entries in Turner, 1966 and Burrow and Emeneau, 1961 and 1968, he demonstrates that comparative analysis of the onomatopoetic forms in a wide range of South Asian languages allows the postulation of at least 40 sets of pan-Indian etymologies. A sample set of areal etyma is given below:

(12) *DED* 1538: Ta. *kuṟukuṟuppu/ai,* 'snoring, stentorous breathing'; Ma. *kuṟukuṟukka* 'to breath with difficulty, make the sound in the throat of a dying person'; Ka. *guṟuguṟu* 'snoring, purring', *guṟa guṟa* 'sound emitted by an angry bandicoot', *guṟṟu guṟṟu* 'growling, snarling, etc.'; Tu. *guruguru* 'snoring, rattling of phlegm in the throat', *gurkugurku* 'roar of a tiger, grunting of a pig'. With formative *-kk-:* Ma. *kuṟukkuka* 'to purr, coo as a dove', *kurkku* 'a snore'; Ka. *guṟake, gūṟuku* 'snoring'; Tu. *gurku* 'roar (of tiger), grunt (of pig)', *gurkāy(i)suni* 'to grunt, growl, snore'; Te. *guraka* 'snoring'. *CDIAL* 4207: N. *gurgurra* 'purring', *guruguru* 'rumbling;' G. *gurgurvũ* 'to growl, rumble'; M. *gurgarṇẽ* (also listed in *DED* as *gurgurṇẽ, gurgurāvṇẽ*) 'to growl, snarl'; with formative *-k-*, S. *gurkaṇu* 'to purr', M. *gurakṇẽ* 'to growl, snarl'. 4486 and 4489: Skt. (medical, *kāvya) ghuraghurāyate* 'gurgles, snorts, wheezes, puffs', *ghurghurikā-* 'gargling', *ghurghurā-* 'growling (of cat or dog)', *ghurghuraka-* 'gurgling'; Pa. *ghurghurāyati* 'snores'; Pkt. *ghurughuraï, ghurahuraï* 'cries out', *ghuruhuraï* 'growls'; P. *ghurghur* 'snarling'; Ku. *ghurghurṇo* 'to snore'; N. *ghurghurāunu* 'to snarl, grunt'; H. *ghurghurānā* id., (Platts) 'to rumble, snort,

> snore', M. *ghurghurṇẽ, ghurghurāvṇẽ* 'to roar, snarl', Si. *guguranavā* 'to thunder'; with formative *-kk-*, Pkt. *ghurukkaï* 'thunders'. 4487, with formative *-k-:* S. *ghurkaṇu* 'to snarl'; H. *ghur/ṛaknā* 'to growl'; G. *ghurakvũ* id. (alongside such forms as N. *ghurnu* 'to snore, coo (of doves)'; Or. *ghuribā* 'to gargle') . . . (Emeneau, 1969:291).

Comparison of the forms cited above allows Emeneau to reconstruct the forms Dr. **guṟuguṟu-, *korakora-* (? ṟ); Indo-Aryan **g/ghurug/ghuru-, *khuru/akhuru/a-.*

In assessing the importance of these onomatopoetics as an areal feature, Emeneau is able to conclude that

> (1) In the language families of India there is a common pattern of onomatopoetics with great proliferation of items in all the languages and some areal etymologies. . . .
> (2) The IA family does not inherit the pattern from IE. . . .
> (3) Consequently we may postulate diffusion of both the pattern and some etymological items from the indigenous families into IA. (Emeneau, 1969:288).

Emeneau's third conclusion if true is highly significant. In spite of the wealth of areal features demonstrated for South Asia, the role of a Dravidian substratum in the history of Indo-Aryan has not always been accepted. Even Jules Bloch, who has contributed as much as any scholar to our knowledge of areal features, downplays the possible importance of Dravidian to the development of Indo-Aryan, stating that:

> En l'état actuel de nos connaissances, rien ne permet d'affirmer que l'aspect pris par l'aryen dans l'Inde soit dû à son adoption par des populations de langue dravidienne. Si le substrat y est pour quelque chose, ce substrat peut au moins également bien se chercher dans d'autres familles, spécialement la famille munda. (Bloch, 1925:20).

To Bloch, proximity with Dravidian merely facilitated linguistic developments whose seed was already present in Indo-Aryan. This was especially held to be the case with regard to the development of a set of retroflex consonants:

> Rien donc ne permet d'assurer que les cérébrales indo-aryennes soient d'origine indigène. La prononciation locale a rendu possible le développement de cette catégorie; et en ce sens l'action du substrat est indéniable. Mais il faut immédiatement insister sur le fait que les langues munda ont des dentales et des cérébrales tout comme le dravidien; et rien n'empêche donc théoriquement d'admettre à l'origine de la prononciation sanskrite l'action d'un substrat munda ou apparenté au munda, sinon d'une

> quatrième famille linguistique inconnue.
> (Bloch, 1925:6)

Emeneau's conclusions with regard to the onomatopoetic forms seem to take care of Bloch's reservations, demonstrating the possibility of direct borrowing of grammatical features from Dravidian into Indo-Aryan, and indicating the possibility of other features being similarly transmitted.

Other areal aspects of onomatopoetic forms have been discussed in Dimock, 1957. In this paper, Dimock points out their existence in Bengali, describes their phonological and morphological properties and briefly examines their history. He tries to show that there are rules of sound symbolism--"phonaesthetics" in Firth's terms--by which the phonological components of onomatopoetic forms are conventionally considered to have semantic properties. This is similar to attempts to demonstrate that there are semantic properties held in common by the *sl-* in such English words as *slide, slippery, sled, slope, slink* (roughly that of slipperiness) or to claim that high front vowels somehow convey smallness or high pitch as opposed to largeness or low pitch for low or back vowels (e.g. cheeping and chirping vs. growling and barking). Dimock tentatively suggests a number of conventional symbolic values for particular Bengali vowel phonemes:

> /ɔ/ 1. "extremity in terms of quantity or quality" (*kɔskɔs* 'extreme heat'), 2. "throbbing, shaking, glittering, or flickering motion of appearance" (*bhɔlbhɔl* '(blood or water) being pumped or flowing in spurts'), 3. "rottenness or softness" (*pɔcpɔc* 'feeling of an over-ripe fruit');
> /æ/ "something decidely unpleasant, either in the nature of the thing indicated, or its effect upon the speaker" (*kæṭkaeṭ* 'shrewish loud-voiced woman');
> /i/ "lightness, and, sometimes, insubstantiality" (*khikkhik* 'a baby's laugh');
> /ã/ "extremity of an unpleasant type" (*gããgãã* 'shouting at the top of one's lungs').
> (Dimock, 1957:28-9).

In addition to suggesting that the term, "phonaestheme," might be relevant for describing the symbolic value of certain Bengali vowel phonemes, Dimock further suggests that given the wide dissemination of onomatopoetic forms in South Asia the use of certain phonaesthemes might be an areal phenomenon.

As interesting as these suggestions might be, they must be rejected at present. The existence of commonly held symbolic values for phonemes such as those proposed seems dubious even in Bengali, and relies heavily on impressionistic judgments derived from a small set of data. The psychological literature in the west, seeking to verify universal or culture-bound symbolic values of particular phonological units has at best been ambiguous. Although speakers in a wide variety of cultures seem to be able to choose between alternate semantic characterizations of nonsense syllables (e.g. which is larger, a *glip* or a *glop*?) speakers cannot associate a recurrent semantic property with a single phoneme (i.e. attribute a particular meaning to the vowels in those words). Moreover, the setting up of such correspondences is often possible only by carefully restricting the data on which it is based, and frequently there are numerous exceptions to any rules postulated. Prudence dictates that Dimock's symbolic values in Bengali require considerable psychological testing before they can be accepted.[9] In the absence of corroborating evidence from other contemporary South Asian languages, it is not possible to justify a commonly held system of sound symbolism as an areal feature. Nor is it feasible to speculate on the origins of such a system, were it demonstrated that one exists.

There are, however, other areal features of greater validity than the last. Bright, in an important paper (Bright, 1966a), discusses the occurrence of a linguistic process of noncontiguous vocalic assimilation or metaphony, by which vowels of adjoining syllables are brought into mutual phonological conformity. The specific rule with which Bright begins his discussion is one from South Dravidian phonology in which the vowels of word-initial open syllables are lowered when followed by syllables containing the low back vowel *a* (e.g. PDr. *iṭ- > PSDr. *eṭ-aṟu 'stumble'; PDr. *pur- > PSDr. *por-ay 'layer'). Bright has four main points to make with regard to this rule:

> (1) This process is reflected in all modern South Dravidian languages, though it is obscured in the literary dialects of Tamil and Malayalam by a change of an opposite, dissimilatory type. (2) The process, which in its earliest form affected only high short vowels, has spread in some languages to affect other vowels, including long ones. (3) This enlarged scope of the process has in some cases produced new phonemic distributions and expanded phonemic inventories. (4) The process has parallels in non-Dravidian languages of South Asia, raising the possibility that a linguistic area or Sprachbund, made up of languages displaying metaphony, may be recognized in the Indian subcontinent (Bright, 1966a:311).[10]

There is somewhat more difficulty in accepting metaphony as a pan-Indian areal trait than in accepting some other features described above. First are problems of distribution. In modern Indo-Aryan the phenomenon is widespread only in Konkani, Bengali, Assamese, and Sinhalese. Bright himself notes that metaphony seems to be of little or no importance in the contemporary Munda languages, and we can find traces of it only in reconstructed early stages of that language family. There is also a qualitative problem. Many of the instances of metaphony cited by Bright differ considerably from one another. The South Dravidian phonological rule discussed earlier lowers an initial high root vowel (i.e. [u] or [i]) in an open syllable when preceding a low back vowel [a] in the following syllable. The data from Indo-Aryan, on the other hand, shows types of metaphony in which the conditioning factor of the vowel change is not the vowel in the following syllable but rather a following consonant cluster, e.g., Skt. *nidrā* 'sleep' < Pkt. *ṇeddā*, *ṇiddā* (Bright, 1966a:320). Conformity of the metaphony with South Dravidian norms seems to decrease as one moves away from the areas of South Dravidian languages. In Munda, in contrast to the South Dravidian pattern, the metaphony occurs progressively as well as regressively. Thus it seems that if metaphony is to be accepted as a synchronic areal trait, it must be with the proviso that it is not a specific rule (i.e. morphophonemic change occurring in a defined context) which is the shared trait. Bright is aware of the limitations in his data and of the difficulties in extrapolating historical generalizations from them.[11]

Any importance of metaphony is not then in its usefulness as a distinctive "cross-language-family" marker of linguistic convergence, but rather as a "conceptual" feature which was transferred in a limited arena of multi-lingual contact, and which subsequently spread out and was adapted according to the different natures of the languages into which it was incorporated. Bright concludes that

> it is possible to hypothesize that each of the non-Dravidian developments represents a period of influence from the Dravidian pattern, followed by a generalization of the pattern to involve such factors as progressive assimilation, raising, fronting, and complete assimilation. By this hypothesis, a linguistic area would indeed be recognized, including most of the eastern and southern parts of the Indian subcontinent. Support for the hypothesis must come from new linguistic data, as well as from material on the nonlinguistic history of India. (Bright, 1966a:322).

Two further important South Asian areal features have been pointed out by Emeneau in his 1974 article "The Indian Linguistic Area Revisited." The first of these is the fact that both Indo-Aryan and Dravidian have morphemes with a particular set of semantic and grammatical properties. While the phonological realizations of the morphemes differ greatly across the language families, their uses are remarkably similar. The second feature involves sets of lexical items in various South Asian languages which are semantically interrelated in common ways and which interrelate with caste structures in ways too similar to be due to chance.

The first areal feature is exemplified by Indo-Aryan _api_ and Dravidian _*-um_. Sanskrit has a form _api_ which was inherited from Indo-European _*epi/*opi_ and which has five distinct functions:

> Usage 1: 'also', i.e. this as well as, in close connection with the previously stated noun, non-finite verb, predication, etc.
>
> Usage 2: 'and'; in series, usually but not always in combination with _ca_, i.e. _cā'pi_, but also the non-enclitic _api ca_, 'both' 'and also'. . . .
>
> Usage 3: in concessive constructions, 'even', either with _yadi_ 'if' (_yady api_ 'even if, even though', often followed by the correlative _tathā'pi_ 'even so'), or following a single word or phrase. In either type of construction the meaning is: 'if X is/was added to the situation, still Y, the opposite of what is to be expected as the result of X, will be/was the case'. . . .
>
> Usage 4: . . .following a numeral or numeral phrase, _api_ indicates that all members of the numbered group, without exception, share in the statement; it is sometimes redundantly used with _sarva-_ 'all,' _sakala-_ and _samasta-_ 'all, the whole'; sometimes it occurs with adverbs such as _sadā_ and _nityam_ 'always'. E.g., _dvāv api_ 'the two of them, both', _sarve'pi_ 'all of them, without exception', _sakalā'pi_ 'the whole of it'. . . .
>
> Usage 5: . . . _api_ with an interrogative pronominal form or derivative produces an indefinite phrase; e.g., _ko 'pi_ 'whoever, someone, anyone'. . . . (Emeneau, 1974a:94).

Usages 1-3 are found as far back as Vedic Sanskrit, while the others are later developments--clearly extensions from the earlier senses of the construction. NIA languages show forms in any or all of the five functions which are either historically derived from _api_ or substitutions of different morphemes for it in the same functions. A tabular summary of the forms used for the five functions of Sanskṛiṭ *api* in four NIA languages is given below:

Usage	1	2	3	4	5
Marathi	hĩ, bī	āṇi, va	-hī, -ī	-hī, -ī -gh-	-hi
Maithili	-o, -hu	o, āo, āor, ār	-o, -ao	-o, -hū	-i, -u
Braj	hū, h, ū̃, ū	aru	?	-hū,-hū̃ -ū, -ū̃	hū, hū̃, -u, -ū
Hindi-Urdu	bhī	aur bhī	bhī̃	-õ	-ī (and other forms)

Figure 51. Five functions of Sanskrit *api* in four NIA languages (from Emeneau, 1974a:96).[12]

This construction has strong parallels in Dravidian, although the Dravidian constructions do not use *api* or any other form historically derivable from it, but rather substitute **-um*. The construction is of great antiquity in Dravidian, and occurs in all five of the uses cited above (although any given Dravidian language may not show all five). From written records it is clear that Proto-Dravidian used this construction at least as early as the emergence of usages 3, 4, and 5 in Indo-Aryan (cf. Emeneau 1974:111). Emeneau concludes that the presence of the parallel constructions can best be explained by diffusion from Dravidian into Indo-Aryan. The difference in phonological shape between Indo-Aryan *api* and Dravidian **um-* is attributed to bilinguals having calqued--substituted a morpheme in one language for a synonymous one in another--the former for the latter. Emeneau speculates that "the uneven disintegration of the structure in NIA is due to a differential extinction of Dravidian in different sub-areas, earlier in the Gangetic valley, where, e.g. the disintegration of normative Hindi has gone deceptively far, later in the area of Marathi, where there is hardly any disintegration of the structure" (Emeneau, 1974a:111).

The other areal feature cited in this article involves lexical items of various South Asian languages. Emeneau notes that virtually all South Asian languages make phonological distinctions between terms designating male and female members of various castes and subcastes (which may, of course, sometimes be indistinguishable from occupational terms) and cites *CDIAL* and *DED* as containing numerous items of this type.[13] Emeneau considers the use of such paired lexical items to be indigenous to India (i.e. pre-Aryan), and to have arisen as part of an historical "Indianization" process in Indo-Aryan. This view implies that Indo-Aryan has witnessed a gradual adaptation of parts of its lexical system, and has become increasingly facile in expressing concepts which are in accord with a non-Indo-Aryan social order. This somewhat Whorfian position considers the linguistic adaption of Indo-Aryan to be part of a more general process of cultural adaption.

Perhaps the most innovative recent work on the Indian linguistic area has been carried on during the past few years by Colin Masica in his Ph.D. dissertation, "A Study of the Distribution of Certain Syntactic and Semantic Features in Relation to the Definability of an Indian Linguistic Area" (Masica, 1971), in an article, "The Basic Order Typology as a Definer of an Indian Linguistic Area" (Masica, 1974), and in a paper written jointly with A. K. Ramanujan, "Toward a Phonological Typology of the Indian Linguistic Area" (Ramanujan and Masica, 1969). Unlike earlier investigations of the South Asian linguistic area, Masica is interested in determining the extent to which the known bundle of shared linguistic features is unique to South Asia. Past literature on *Sprachbund* phenomena has been confined to demonstrating shared linguistic features and to describing the processes of convergence which led to this sharing. Masica's concern is rather with the extent to which these features are purely South Asian ones, and with whether such a linguistic area might be part of some larger "Asian *Sprachbund*."

The linguistic traits which Masica examines in dealing with these questions are not all the same as the features which have traditionally been examined in the Indian linguistic area literature. Rather, he examines a number of variables involving word order which Joseph Greenberg has advocated as basic indicators of linguistic typology,[14] such as the relative positioning of subjects, verbs, and objects in surface structures. Some languages of the world regularly place the verb at the end of clauses and after subjects and objects, while other languages generally place the verb between the subject and the object. In his 1974 paper, Masica examines this and other syntactic variables in five Indian languages (Hindi-Urdu, Bengali, Telugu, Malayalam, and Santali), and compares the results obtained with a large sampling of non-South Asian languages extending from Europe to East Asia, marking the limits of the range of these syntactic features with isoglosses as in classical dialect geography. If an Indian linguistic area can be said to exist, one would expect it to be surrounded by a thick bundle of isoglosses referring to the traits plotted.

The syntactic traits ostensibly shared by much of the Indian *Sprachbund* are as follows:

1. The presence of verbs which follow their objects rather than precede them. When subjects are present, they will generally precede the objects. (E.g., H. *raaman hindii boltaa hai* 'Raman speaks Hindi'; Te. *kamala puulu koostunnadi* 'Kamala is plucking flowers'; Sa. *uni hoR ko tOlkedea* 'They bound that man'.)

2. The placing in object position of goals of verbs of motion, predicate nominals and adjectives, adverbial complements, and infinitival complements of catenatives. (E.g. H. *sab log apne apne ghar jaao* 'Everybody go home'; B. *tini Ekjon Daktar chilen* 'He was a doctor'; Te. *ciire nallagaa undi* 'The saree is black'; M. *rakSappeTaan nookkukayaayirunnu* 'He was trying to escape'; Sa. *ar ADi OkOc'ak'ko cet'ako e pOrtOnket'a* 'And he began to teach them many things'.)

3. The use of verbal auxiliaries which follow rather than precede the main verbs with which they are used. (E.g. H. *aa rahaa hai* 'He is coming'; B. *jabo* 'I shall go'; LTe. *tswādiwi unTini* 'I had read'; Sa. *gOc' adadea* 'have killed'.)

4. The marking of the syntactic functions of nominal elements by means of postpositions, as opposed to prepositions. (E.g. H. *ghar me'* 'in the house'; B. *boner mOddhe* 'in the middle of the forest'; Ma. *muriyil* 'in the room'; Sa. *bir sEnre* 'towards the forest'.)

5. Patterns of word order in which adjectives, genitive phrases, demonstratives, and numerals generally precede the nouns which they modify. (E.g. H. *andherii raat* 'dark night'; Ma. *kaakkayuTe muTTa* 'the crow's egg'; B. *ei lokTi* 'this man'; Ma. *naalpatu kuzhi* 'forty holes'.)

6. Word order in which qualifiers precede adjectives. Included under "qualifiers" are intensifiers--'very', 'somewhat', etc.--and what Masica calls "qualifiers of comparison"--e.g. greener *than a leaf*. (E.g. H. *bahut acchaa* 'very good'; B. *khub gOrom* 'very hot'; Te. *tsaalaa ettugaa undi* 'it is very high'; Sa. *inren khOne maraṅa* '(it is) bigger than mine'.)
(Examples taken from Masica, 1974:159-64).

Masica considers the presence of these features to collectively constitute an Indian syntactic norm. By comparing the dissemination of these features with their distribution in a sampling of adjoining languages, he is able to conclude that there is indeed such a thing as an Indian linguistic area. He cites four specific ways in which such an area can be substantiated:

> A. First, the basic OV isogloss defines a massive middle segment of the world, a largely left-branching syntactic zone in which India is the main southern anchor. In keeping with its basically north-south trend, we might name this macrozone INDO-ALTAIC. . . .
>
> B. Subsequent isoglosses, especially those involving the structure of the noun phrase, cut the OV belt almost in two. . . and isolate the subcontinent as a separate sub-zone.
>
> C. These subsequent isoglosses. . .define three transitional subzones, where the basic OV syntax begins to give way to opposing phenomena: the Iranian, the Tibeto-Burman, and the Abyssinian. . . .
>
> D. A thick bundle of isoglosses separates India from Southeast Asia beyond Burma, from Arabic, and from Africa beyond Ethiopia. (Masica, 1974:172).

Masica concludes that the area so defined is "not an area of transition (formed by the intersection of isoglosses), but a trait-core area, surrounded by *concentric* isoglosses" (Masica, 1974:172).

In his 1971 dissertation, Masica expands the inventory of features whose distribution is considered in this manner. He compares the spread in India of morphologically marked causative verbs, conjunctive participles, explicator compound verbs,[15] the so-called "dative" construction,[16] and the presence or absence of a morphologically marked verb "to have."[17] The occurrence of these linguistic features in particular regions can be marked on maps, and isoglosses marking the limits of their distribution can be drawn. The isoglossal lines drawn for each of these traits can then be examined for bundling. A map of this type--in which isoglosses marking the distribution of causal constructions, adjective-preceding-noun word order, past gerunds, explicator-compound verbs, dative subject construction, and OV word order are brought together--is shown in Figure 52. This map shows a clear bundling of features in South Asia, and supports claims for the existence of a South Asian *Sprachbund*.

As an independent means of measuring the cumulative importance of the linguistic features which he discusses, Masica assigns an arbitrary numerical weight to each of his variables. By assigning points for each

"Indian" trait shown by a language, he seeks a measure of the "Indianness" of a wide variety of languages on a single linear scale. Thus, for example, a language is given "1/2 point for morphological causatives regularly derived from intransitives and semi-transitives, 1 point for these + derivation from transitives, and 2 points for double causatives (causatives derived from causatives)" (Masica, 1971:218). With the maximum number of points which any language can receive being 20-1/2, values recorded by Masica range from 20-1/2 for Hindi-Urdu to 1/2 for Cambodian.[18] The highest scoring languages on this scale (19-20) are called "typical" Indian languages, and are immediately followed by other Indian languages and Altaic languages (16-18), then Tibeto-Burman, Ethiopic, Georgian Tajik, and Finno-Ugric (12-15). A surprising feature of this ranking is the obvious typological connection between Indian and Altaic languages, which is closer than that found between languages in the geographically more proximate Indian and Tibeto-Burman areas.

Another imaginative approach towards the South Asian *Sprachbund* has been adopted by Masica and Ramanujan. In their 1969 paper, they examine the distribution of phonological oppositions in South Asia and adjoining areas. Operating within a framework of early generative phonology, they seek to identify in a large number of languages the existence of given phonological oppositions and to determine the class of sounds to which these oppositions apply. Once such a determination has been made, these languages can be grouped together on the basis of their sharing a particular phonological opposition or set of oppositions.

Ramanujan and Masica examine consonant systems with reference to the presence or absence of the following oppositions: 1. grave/acute; 2. compact/diffuse; 3. strident/mellow; 4. retracted/unretracted; 5. flat/plain; 6. tense/lax; 7. voiced/voiceless; 8. checked/unchecked; 9. sharp/plain; 10. ingressive/egressive; 11. tone.[19] (Similar oppositions are considered for vowels.) All languages within South Asia exhibit the contrasts indicated by features 1 and 2; features 3-7 are major typological markers and identify important cleavages within South Asian languages; while features 8-11 are diagnostic only for a small number of South Asian languages.

Frequently the distribution of a particular feature may be a strong indicator for typological grouping of languages. For example, the retracted/unretracted opposition can be used to differentiate dental and retroflex apical consonants. This opposition is common among stops (and was in fact one of the early defining criteria of the Indian linguistic area), but is more sporadic with strident affricates. According to Ramanujan and Masica, the latter use of the opposition occurs in loan words "in a belt of languages stretching from Konkani, Marathi and North Kannada to Telugu, Gadba-Ollari and Southern Oriya" (Ramanujan and Masica, 1969:564), while in the South, Toda, and in the northwest, Bhalesi, Burushaski, Ishkashimi, Kafiri, Kashmiri, Khowar, and Pashto make full use of this opposition. We can thus establish geographical areas which differ significantly with regard to this feature if we consider not merely its presence or absence, but the class of sounds to which it pertains.

A major benefit of Ramanujan and Masica's approach to phonological features is that its adoption allows the enumeration of numerous linguistic areas with regard to various phonological features. Situations in which isoglosses form thick bundles may be rarer than once thought. Language can be thought of as standing in an n-dimensional grid of linguistic features, and forming a part of different sets of languages with regard to the different features. The adoption of "abstract" phonological oppositions as criteria for typological comparison provides a natural means for formalizing these intersecting relationships. Once oppositions can be isolated, it makes sense to speak of micro-linguistic areas which can be formalized in terms of them. The approach is limited only by the care with which the variables which are to serve as the bases for comparison are chosen.

In Ramanujan and Masica's case, the selection of retracted/unretracted is extremely productive for typological purposes. It clearly is useful in setting up a standard of "Indianness" against which other languages can be measured. The acute/grave opposition is much less valuable for typology, since the opposition is universal to the languages of the world. Hopefully, in the future it will prove possible to extend Ramanujan and Masica's typological approach beyond phonology to include morphological and morphophonemic alternations and possibly to incorporate notions of markedness.

4.2. Micro-Linguistic Area Phenomena

4.2.0. The types of linguistic contact which existed prehistorically between Indo-Aryan and Dravidian speakers have been going on to a greater

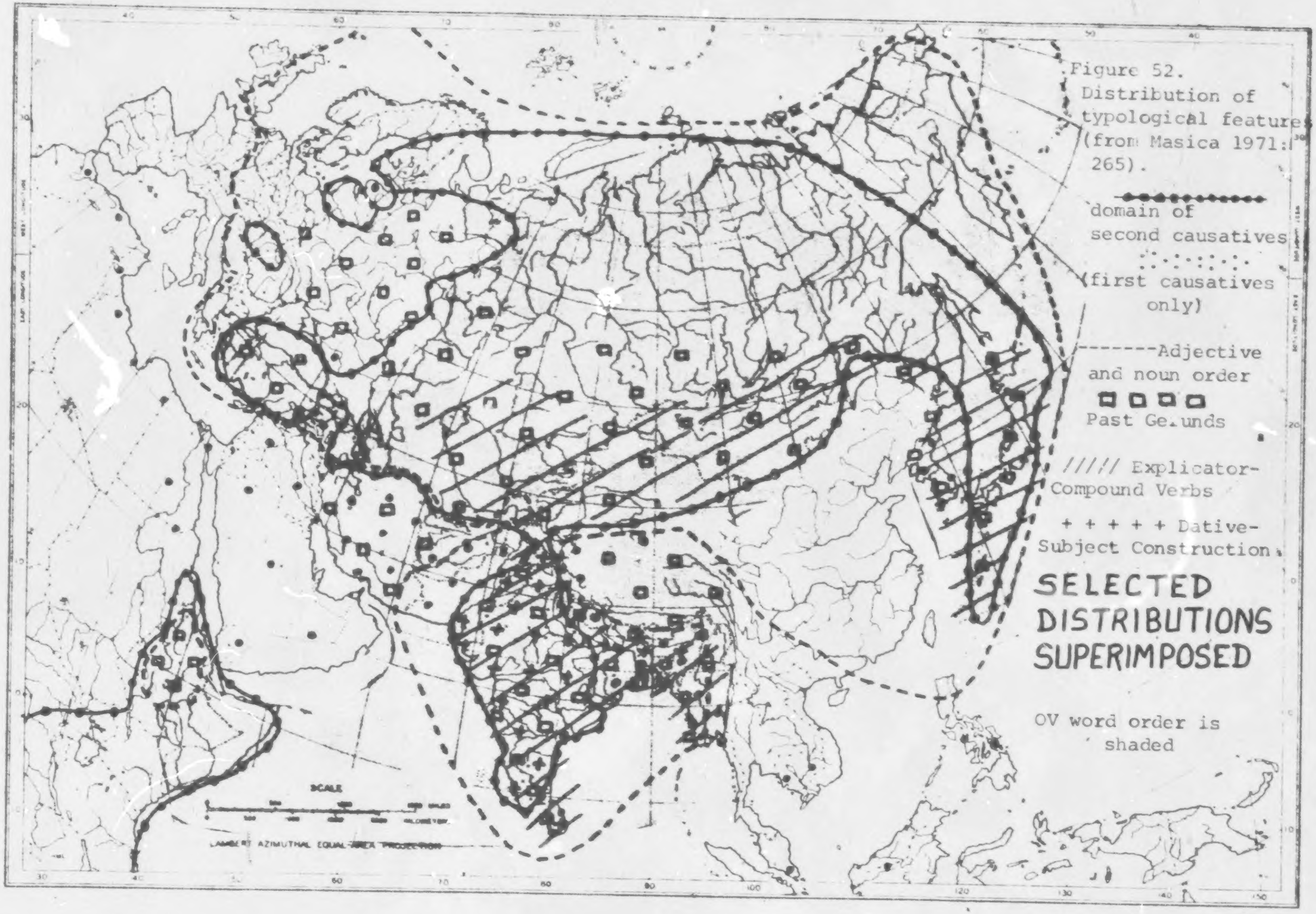


Figure 52. Distribution of typological features (from Masica 1971: 265).

domain of second causatives
(first causatives only)
Adjective and noun order
Past Gerunds
///// Explicator-Compound Verbs
+ + + + + Dative-Subject Construction.
SELECTED DISTRIBUTIONS SUPERIMPOSED
OV word order is shaded
SCALE
LAMBERT AZIMUTHAL EQUAL-AREA PROJECTION

or lesser extent up to the present. It is not unreasonable to surmise that these contacts may have given rise to regions which fit Emeneau's definition of a linguistic area. The literature which pertains to these matters is limited, and is often not addressed directly to questions of this sort.

There are a number of phenomena which clearly should be sought in identifying small-scale linguistic areas. Such an area should include speakers of genetically unrelated languages living in proximity over a protracted period of time, where there are regular social contacts between the speakers of the different languages, possibly leading to widespread bilingualism and/or codeswitching. In addition, there must be evidence of linguistic borrowing between the adjoining codes. In South Asia, a number of areas with such a concentration of circumstances exist: the Himalayas, particularly in areas of contact between Indo-Aryan and Tibeto-Burman; Sri Lanka, with Tamil-Sinhalese contact; the Northwest Indian Frontier with regard to Indo-Aryan, Iranian, and Dardic; and Dravidian-Indo-Aryan border areas. We discuss briefly the first three of these here.

4.2.1. Nepal

The only real evidence we know of which directly treats the possible existence of a Himalayan linguistic area is found in Bendix, 1974. In this article he examines Nepali, an Indo-Aryan language and the official language of Nepal, and Newari, a Tibeto-Burman language spoken largely in the Kathmandu valley. The languages have been in contact for centuries and, according to Bendix, Indo-Aryan speakers from north India have repeatedly been assimilated to Newari. He tries to show, albeit inconclusively, that although the formal shape of various verbal morphemes varies considerably between Nepali and Newari, the paradigmatic relations between the abstract syntacto-semantic categories which these morphemes represent is relatively similar. In other words, Nepali and Newari operate with essentially similar verbal categories, but express them with different forms. Bendix feels that there is a fundamental similarity in the verbal categories underlying the Nepali and Newari verbal alternations shown in Figure 53. He is able to derive from these forms a list of what he considers to be "approximate tense aspect correspondences" (Figure 54).

Another linguistic feature cited by Bendix as being of possible areal

Nepali

(1) perfective	
a. *gar-yo*	'did'
b. *gar-eko*	'done'
c. *gar-eko cha*	'has done'
d. *gar-eko thiyo*	'had done, did'
(2) imperfective	
a. *gar-ne*	'doing'
b. *gar-cha*	'does, will do'
c. *gar-thyo*	'used to do, would have done'
d. *gar-la*	'will/might do'

Newari

a. *yat-o*	'did'
b. *ya-v'*	'does/do, used to do, did'
c. *yan-a*	'(I/we) do, used to do, did'
d. *ya-i*	'will/might do'
e. *ya-e*	'(I/we) will do'
f. *ya.v'-gu, yan-a-gu,* *yai-i-gu, ya-e-gu*	'having done, doing, going to do, etc.'

Figure 53. Nepali and Newari verb tense forms (from Bendix, 1974:44).

Newari	Nepali	Glosses with 'do'
-gu	*-ne*	'doing'
	-eko	'having done'
-gu du	*-eko cha*	'has done'
	-eko thiyo	'had done'
-o	*-yo*	'did'
-v'	*-thyo*	'used to do'
	-cha	'does'
-i		'will do'
	-lā	'may/might do'

Figure 54. Approximate tense-aspect correspondences in Newari and Nepali (from Bendix, 1974:54).

significance in Nepal is the use of the so-called "ergative" construction. In many Indo-Aryan languages the subjects of transitive verbs are marked with a special "ergative" or "agentive" postposition if the main verb occurs in some form of the perfective.[20] This construction is historically a reflex of an OIA one in which the verb is a perfective participle and its subject is placed in the instrumental (e.g. "the work was done by him"). In NIA the verb generally does not agree in number, person, or gender with the subject, but rather with the object (e.g. H. *mãĩ ne vah* *kitāb kharidī* (I (erg.) that book (fem., sing.) bought (fem., sing.)). In Nepali, however, in spite of the use of the ergative postposition *le* in these constructions, the verb agrees with its subject. Newari and other adjoining Tibeto-Burman languages are lacking in the ergative construction. Bendix speculates that the presence of an ergative postposition without object-agreement in Nepali represents a compromise between the full Indo-Aryan construction and its complete lack in Tibeto-Burman.

4.2.2. Sri Lanka

The literature which points to the existence of a Sri Lanka linguistic area is relatively insubstantial. The only systematic explication of the subject to date is in a paper by T. Elizarenkova, "Influence of Dravidian Phonological System on Sinhalese" (Elizarenkova, 1972). This paper compares the phonemic systems of Tamil and Sinhalese and tries to isolate a number of phonological features in the latter which appear to have arisen through phonological interference from Dravidian. There are a number of such features pointed out by Elizarenkova, the most important being: 1) the difference in the total number of phonemes in Sinhalese and in the rest of Indo-Aryan with Sinhalese having no more than 30--considerably less than the rest of Indo-Aryan, but more than Tamil); 2) the loss of aspiration in Sinhalese, a feature which remains widespread in Indo-Aryan, but is only marginal in Dravidian (having been introduced into NDr. comparatively recently through Sanskritic lexical borrowings); 3) the partial neutralization of the distinction between *s* and *h* in Sinhalese as a result of the rule *s* → *h* which had been operative as far back as Sinhalese Prakrit; 4) the absence of nasal vocalic phonemes in Sinhalese, widespread in the rest of Indo-Aryan; 5) the opposition of long vs. short vowels, common in Dravidian, but sporadic in Indo-Aryan; 6) the absence of diphthongs in Sinhalese, distinguishing Sinhalese from the eastern languages of Indo-Aryan.

In assessing the importance of these comparative phonological features, Elizarenkova asserts the existence of

> paradigmatic interference of the Tamil phonological system towards the Sinhalese one. . .[which] manifests itself mainly either in the loss of some oppositions of distinctive features in Sinhalese caused by the contact with Tamil (such as the opposition of aspiration of the consonants and that of nasalization of the vowels), or in the change of the

> volume of a certain opposition which has existed in Sinhalese previously. (Elizarenkova, 1972:133).

Elizarenkova's conclusions cannot be fully accepted at the present time. She based her comparisons on standard literary Tamil with modern Sinhalese, while it would appear that any meaningful conclusions on the existence of a Sri Lanka linguistic area necessitate extensive descriptive data on Sri Lanka rather than Indian Tamil. Such literature is sparse, and has only started to appear recently.[21] Moreover, much of Elizarenkova's material is of limited typological value. Although it is interesting that Sinhalese has fewer phonemes than the rest of Indo-Aryan, the significance of the fact is limited by the lack of an explicit theory of phonology which explains the distribution of phonemes in terms of configurations of abstract distinguishing features, be they "Pragueian" or transformational-generative. In other words, numbers of phonemes per se tell us nothing about the systems of which these phonemes are a part, and Elizarenkova only hints at the properties of such systems. It would also be useful to know the extent to which the features observed in both Sri Lanka Tamil and Sinhalese are restricted to Sri Lanka. Might they not be part of a general South Indian linguistic area, in which case the presence of non-Indo-Aryan linguistic features in Sinhalese would not necessarily be attributable to Tamil influence? Elizarenkova has provided a basis for carrying on an investigation of such questions, although much of the actual work is yet to be done. Furthermore, a meaningful *Sprachbund* must contain a wide variety of shared linguistic properties, including lexical, morphological, syntactic, and possibly semantic aspects. Until such evidence is forthcoming from Sri Lanka data, we cannot draw any firm conclusions about the existence of a linguistic area. (The same cautions are equally true with regard to the Nepal area or any other proposed *Sprachbund.)*

4.2.3. Northwest Indian Frontier

This last micro-linguistic area involves a type of linguistic convergence qualitatively different from the sort discussed previously. It is well known that the Mughul conquest of the Indian sub-continent had a profound effect on many of the indigenous languages of the region. Perso-Arabic phonemes, lexical items, morphemes and, on rare occasions, syntactic constructions were introduced into numerous Indo-Aryan languages.[22] These borrowed forms include some that have pervaded every linguistic stratum in North India and some that occur only in highly restricted (and often formal) contexts.

The pervasiveness of Perso-Arabic elements in the Indo-Aryan languages of north India has been advanced by education, particularly of males, in Urdu, and by the popularity of Urdu novels and films throughout much of India. Film songs, heavily Perso-Arabicized, are listened to throughout the subcontinent. In spite of their widespread acceptance, in no case have Perso-Arabic elements been consistently absorbed at all styles and levels of use within a particular language. The use of Perso-Arabic elements is socially conditioned and may be correlated with a number of social variables, degree of education, religion, formality of speech contact, etc.

The acceptance of such sets of correlations is not purely geographical, and transcends linguistic boundaries. We are dealing with a situation in which a number of codes have numerous socially conditioned varieties, and in which some of these varieties involve the adoption of Perso-Arabic normative forms. Speakers of a number of languages thus share similar patterns of variability in their language use, agreeing on the use of particular linguistic elements in a similar range of constructions. The absolute frequency of use of these constructions may vary from one geographical region to another, or vary with the education of the speaker, but the overall contour of the stratificational pattern will stay relatively constant. This can possibly be looked upon as an areal trait.

A related set of linguistic traits may possibly be of areal significance in the northwestern part of the subcontinent. Many of the northwest Indo-Aryan languages (e.g. Hindi, Rajasthani, Punjabi, Gujarati), Dardic languages, and eastern Iranian languages show a considerable amount of linguistic convergence. The most important of these shared traits are the pervasiveness of Perso-Arabic phonemes and lexical items, the borrowing of Indo-Aryan lexical items into the non-Indo-Aryan languages, and the assimilation in at least one case (Pashto) of aspirate consonants into Iranian. In the case of border languages spoken in post-1947 India, the presence of Sanskritization in at least some strata of the languages may serve as an areal trait.[23] Once again, however, it is necessary to assert that the northwestern frontier has not been subjected to the same sort of areal analysis as India on a whole, and that it is premature to issue any statements about the existence of a linguistic area.

4.3. Historical Implications of a South Asian Linguistic Area

The material concerning the Indian linguistic area is interesting from a number of points of view. When comparing similarities in linguistic structure of a number of languages cutting across genetic stocks, the concept of "linguistic area" allows us to express generalities of linguistic structure complementary to those expressable in traditional comparative-historical terms. The usefulness of these areal generalizations is further increased if they can be tied in with historical explanations as to how the particular language convergences came about. In which direction and in what order did the borrowings which lead to convergence happen? More profoundly, under what conditions did they occur, and what motivated them? Certainly none of these questions is answerable at present; and for the latter two, we have only rudimentary theories available to deal with them.

There is some literature which can bear on the discussion of these questions. One early attempt to discuss the genesis of the South Asian linguistic area is an article by Emeneau, "Linguistic Prehistory of India" (Emeneau, 1954), which gathers together an inventory of areas of Dravidian influence on Indo-Aryan. Although the main concern in the article is to rectify the failure of scholars to recognize this Dravidian element in Indo-Aryan, Emeneau states that the borrowing of linguistic features was a necessary consequence of the contact between linguistic communities, and that bilingualism enabled the transmission of features from one language to another.

The linguistic prehistory of the region is also treated by Chatterji in a Presidential Address to the All India Oriental Conference (Chatterji, 1953), in which he outlines a number of instances of racial and cultural intermixture in India which he sees as leading to "the Indian synthesis" and the creation of "the Indian man." The article lists a number of features shared across language families in India (concentrating on areas of Dravidian influence on Indo-Aryan), and also speculates on the social and/or linguistic factors which led to the virtual dying out of the non-Indo-Aryan languages of North India, including:

> 1. the prestige of the Aryan speech as that of a *Herrenvolk* which had established itself in the country, and to which the allegiance of the conquered peoples was a matter of course;

> 2. absence of cohesion among the polyglot non-Aryans of Dravidian, Austric and Kirāta origin, living side by side, with the Aryan speech coming to the forefront as a very convenient *lingua franca;*
>
> 3. the spirit of *laissez faire* and an evident policy of non-intervention with reference to the non-Aryan languages--nobody ever seems to have tried to put a stop to or restrict their use. . .;
>
> 4. the liberal policy shown, doubtless as a matter of convenience, by Brāhmans and other custodians of the Aryan's language towards non-Aryan vocables and idioms,--the gradual and unrestricted entry, mostly by the back door, of a large non-Aryan vocabulary first in Vedic and in the Prakrits and then in the Classical Sanskrit, took away the edge of opposition to Sanskrit and other forms of Aryan, if there was any such opposition at all: the gradual approximation of Sanskrit and the Prakrits to the spirit of both Dravidian and Austric made the Aryan's language easily acceptable to non-Aryan speakers;
>
> 5. the fact that Sanskrit and other Aryan [languages] became the vehicle of a great composite culture, all-inclusive in scope, that was being built up through the combined efforts of Ārya, Drāviḍa, Nishāda and Kirāta, helped to maintain its supreme position in a new Indian population of mixed origin, directed more or less by groups like the Brāhmans boasting of a pure Aryan tradition;
>
> 6. the early development of a literature in Sanskrit through the collection of Vedic Hymns and sacrificial texts, and through the redaction of masses of national legendary and semi-historical tales and traditions as in the Purāṇas, gave to Sanskrit an immense advantage over other languages. . .;
>
> 7. it is exceedingly likely that there was no effective linguistic or cultural patriotism (if there was any at all) among the leaders of the various non-Aryan groups in Northern India: particularly when the Brāhmans through their intelligence and prestige were able to give a theory of society which ignored the racial and linguistic aspects and included the whole of Indian humanity within a single scheme. . .;
>
> 8. the inherent beauty and force of the Aryan language which was something which fulfilled the intellectual requirements of the Indian Man, satisfied his aesthetic sense, and at the same time was not foreign to his mental atmosphere if he still spoke or lived in the atmosphere of a non-Aryan tongue. (Chatterji, 1953:50-1).

The eight features spelled out above are of unequal validity, and numbers 7 and 8 are little more than nationalistic rhetoric. Feature 1 is not an explanation of the phenomena being treated, but rather a rephrasing of it. The prestige of Indo-Aryan is moreover not something which can be assumed. The use of the term "prestige" implies a ranking of subjective attitudes towards a number of alternative speech codes available for use at a given time, and these attitudes must be kept apart from the objective patterns of variation in the use of alternate codes. Variation in linguistic performance is often correlated with any of a number of social, economic, and linguistic variables (cf. Chapter 5), and this sort of variation is not necessarily a direct function of the subjective evaluations of speakers concerning the relative prestige of the codes.

Chatterji's point number 2 also explains less than it might appear to at first glance. When talking of the lack of cohesion among the "polyglot non-Aryans of Dravidian, Austric, and Kirāta," he is, in effect, explaining the cultural victory of one of four juxtaposed linguistic families in terms which become meaningful only if that linguistic family has already become victorious. It is equally valid to state that Indo-Aryan, Munda, and Tibeto-Burman displayed a lack of cohesion in contrast with Dravidian. If Indo-Aryan gained relative strength because of "convenience" as a *lingua franca* we need to know why. What were the social, economic, and linguistic forces which necessitated or facilitated the use of Indo-Aryan as a *lingua franca?* Who used it as such, and for what purpose? And is it really the case that Indo-Aryan was used as a *lingua franca* to facilitate inter-group communication? Recent research casts considerable doubt on this last point,[24] and it is not now possible to assume that Indo-Aryan served functions such as those stated by Chatterji.

Perhaps more importantly, the thrust of Chatterji's argument is that Indo-Aryan, through the inherent worth of its culture and literature, and the tolerance of its users towards non-Aryan tongues, naturally became the prestige and accepted idiom in north India. This was facilitated by the lack of organized resistance on the part of non-Indo-Aryan and by the cultural diffuseness of the other language families, which enabled Indo-Aryan to serve a unifying function. These statements must be viewed as biased speculation which explains nothing.

It must be kept in mind that languages are used by individuals and groups in particular social contexts. Communication takes place for a multitude of purposes. Options are always open concerning the use of particular idioms, lexical items, stylistic devices, and the actual codes of expression. Speakers individually display different attitudes toward the various options open to them, and they may or may not be consciously aware of these attitudes. Collectively, speakers establish norms about the utilization of certain options in particular contexts and for particular purposes. The inacceptability of many of Chatterji's statements stems from his failure to take into account the variety of options open to individuals within areas of multi-lingual convergence. Before we can accept conclusions such as Chatterji's, we would need to know what the nature of the contacts were between Indo-Aryan speakers and speakers of other languages. What linguistic options were open to individuals involved in such contacts? What influence did the usage of individuals in such linguistically complex situations ultimately have on the structure of the normative codes which later evolved (all of the ancient Indian codes of which we have written records are in fact normative codes)? What was the relationship between these normative codes and the codes which individuals were using, either in relatively homogeneous linguistic groups, or in contact with members of other linguistic groups?

These questions are difficult ones, and their solution requires considerably more information about language use in prehistoric India than is available to us. Unfortunately, conclusions about language use in prehistoric India can only be based on extrapolation from current patterns of language use, particularly in border areas of multi-lingual convergence, and from the study of the history of languages which are spoken in such areas of convergence.

Kuiper in "Genesis of a Linguistic Area" (Kuiper, 1967) examines the available evidence which can shed light on the origin of the South Asian linguistic area. More particularly, he attempts to trace as far back as possible three major areal traits, in order to determine the patterns of borrowing at early stages of contact between the various language families in the area. Kuiper is partially motivated by a desire to respond to those who stated that many areal features found in Indo-Aryan were not due to the influence of indigenous languages (i.e., Munda and Dravidian) but were the result of processes already operative in Indo-Aryan. The three linguistic features which Kuiper treats are the presence of a series of retroflex consonants, use of a quotative element (such as Sanskrit *iti*), and the use of a gerund construction.

The non-Indo-Aryan origin of retroflex consonants in Sanskrit had specifically been denied by some authors and qualified to the point of triviality by others.[25] While many had pointed out the presence of retroflex consonants in Dravidian as a likely source of their existence in Indo-Aryan[26] few if any had looked at the phenomenon from an historical point of view. Kuiper examines Vedic Sanskrit in some detail to determine precisely the contexts in which retroflexes were first used. He notes that they are of limited frequency in Ṛgvedic Sanskrit, and that many of the words in which they occur are possibly of foreign origin; the frequency of occurrence increases, however, in the Atharvaveda.

The total data which Kuiper cites lend credence to the view that retroflexion in Indo-Aryan, if not completely attributable to Dravidian influence, at least resulted because the borrowing of lexical items from Dravidian triggered a rearrangement of Indo-Aryan phonological structure.

> It may seem natural, then, to assume that. . .in prehistoric Indo-Aryan bilingual speakers who recognized a phonemic contrast between dentals and retroflexes in the foreign language, came to interpret the allophones of proto-Indo-Aryan in terms of the foreign phonemic system. The loan words with retroflexes which. . .they must have introduced into Indo-Aryan may have contributed considerably to the spread of this novel phonemic distinction among the speakers of early Indo-Aryan. (Kuiper, 1967:144).

Kuiper points out that such borrowing must have occurred in a complex social environment. He feels that Dravidianization of the Indo-Aryan phonological system did not occur at a uniform rate throughout Indo-Aryan society, asserting that there was probably a reluctance of higher class Indo-Aryan groups to adopt Dravidian lexical items containing retroflexes unless those items had previously undergone Sanskritization.[27]

The use of a quotative form *iti* in Sanskrit, unlike the presence of a series of retroflex consonants, had antecedents in Indo-European outside of Indo-Aryan, being cognate with Avestan *uiti*. The form as used in Avestan always introduces a following quotation; in Ṛgvedic, however, some instances of this form follow the quoted material, and in later Sanskrit this post-quotative use of *iti* becomes standard. The latter form of the construction corresponds precisely to the common Dravidian pattern (with the substitution of a Dravidian morph for Sanskrit *iti*).

The gradual accomodation of the quotative construction in Indo-Aryan to the normative Dravidian order, as well as the total lack of this post-quotative ordering in Avestan, is an indication to Kuiper that the inherited construction began to be adapted to the Dravidian pattern some time prior to the Ṛgveda. The construction was gradually brought into complete conformity with the Dravidian norm, and subsequently spread to the point where it achieved consideration as an areal feature.

The third feature described by Kuiper, that of the innovation of gerund constructions in Indo-Aryan (and in some Munda languages) which are directly parallel to those found in Dravidian, is interesting especially in its historical aspects. The occurrence of these constructions is minimal in early Ṛgveda, but increases considerably by the end of those texts. However, the morphological makeup of the recent Indo-Aryan constructions seems to be based on ancient instrumental forms which were lost long before the time of even the earliest Ṛgveda texts. This apparent historical anomaly--the borrowing of a syntactic construction at a very early stage, but only much later incorporation into religious poetry--is accounted for by Kuiper

> by assuming that long before the oldest hymns were composed the use of gerunds in proto-Indo-Aryan arose among bilinguals, presumably in colloquial speech. The poets continued to use the traditional perfect participle until, in the last period of Rigvedic poetry, when the influence of other social classes became stronger, the new formation was fully accepted even in poetry. (Kuiper, 1967:152).

4.4. Mechanisms for Genesis of a South Asian Linguistic Area

Franklin Southworth has also addressed questions concerning the origin of the South Asian linguistic area in a number of recent works (Southworth, 1971, 1974; Southworth and Apte, 1974b). In one paper (Southworth, 1974), he attempts to determine the social contexts in which linguistic convergence took place, and to describe the mechanisms of feature transmission. Although his explanations are highly speculative, and their verification awaits considerably more research on language stratification in ancient India, they are interesting attempts to put linguistic borrowing of grammatical features into a social context. Because of their interests we quote from them here at some length.

> The evidence seems sufficient to postulate that at some point the early IA speakers constituted a small but dominant

> minority in a limited area of the northwestern part of the subcontinent. They were probably integrated gradually into the existing social structure, primarily (if not exclusively) at the top levels. Subsequently, a much-modified form of the original IA language became the prestige language of the northwest, spreading gradually southward and eastward as a military and trade language. While the population in the western areas (present day Maharashtra, Gujarat, Sindh) was probably mainly Dravidian-speaking, in the Gangetic plain (especially from eastern Bihar eastward) the IA language was taken up by a predominantly Tibeto-Burman-speaking population. At some point, intellectuals in the upper strata probably became aware of the drastic changes which the language was undergoing, and founded the science of grammar partly to prevent further decay. It is conceivable that this development coincided with the move into the Gangetic region. All of this must have happened before the third century B.C. when we have the evidence of the Ashokan inscriptions. . .to tell us of the extent of the differences between literary Sanskrit and the contemporary administrative language.
>
> The subsequent linguistic history suggests the establishment of political units dominated by IA-speaking elite minorities throughout the present area of speech, followed by gradual adoption of the local IA variety by non-IA speakers. This development may have been preceded and accompanied, in each area, by a process of internal convergence between the relatively conservative form of IA spoken by the elite group, and the highly pidginized variety adopted originally by the non-IA speakers. (Southworth, 1974:222).

In another major paper Southworth attempts to develop the role of pidginization and/or creolization in the transmission of linguistic features (Southworth, 1971). Although we shall come back to the matter in greater detail in section 6.2, it seems fair to state here that it is very likely that cross-language contact areas in ancient India had individuals or communities displaying some form of bi- or multi-lingualism—not of a sort in which the speakers contain full control of fully autonomous codes, but instead have expanded linguistic repertoires enabling them to communicate at least minimally across cultural and linguistic boundaries. The sorts of situations described by Gumperz and others in Delhi vis à vis Punjabi-Hindi bilinguals (Gumperz, 1964), and for Urdu-Marathi-Kannada convergence near the Mysore-Maharashtra border (Gumperz and Wilson, 1971), are probably closer to the situations which must have existed at that time.

It is also likely that pidginization and/or creolization played a major role in the formation of these expanded repertoires. This is not to suggest that all languages in the South Asian linguistic area are necessary "pidgins" or "creoles," but rather that individuals living in areas of multi-lingual convergence needed to communicate within a sociolinguistically complex community. In order to accomplish this, they employed extended linguistic repertoires which accomodated new linguistic features through what are essentially processes of pidginization or creolization. The existence of large numbers of individuals with such expanded linguistic skills would then facilitate the transmission of areal features. A linguistic feature (such as a retroflex consonant in loan words) which was introduced into an individual's competence through some sort of pidginization would become part of his linguistic competence, and possibly result in a fundamental phonological realignment of elements in his native code. Kuiper's data on the introduction of retroflex consonants in Vedic Sanskrit is fully in accord with such a theory.

Southworth has presented evidence concerning processes of this sort in the history of Marathi. It is his contention that modern Marathi is the result of historical fusion of Dravidian and Indo-Aryan elements, and that pidginization and creolization were actively involved in the fusion. If Southworth is correct, we must accord dynamic processes such as pidginization and creolization a greater degree of significance in the history of South Asian languages than has been done previously. It will not suffice to conceive of a South Asian linguistic area in terms of relatively static families of languages which are influenced by the borrowing of linguistic feauures from the languages with which they come in contact. We must rather come up with models of linguistic prehistory which, in essence, bring the convergence into the core of the description.

It is not an historical accident that linguistic features have been borrowed between language families. In situations where communication across language boundaries was necessary, speakers were forced to expand their linguistic skills. With an expanded social range of contacts must have come an awareness of new sets of linguistic prestige values, and the restriction of certain linguistic features to defined social contexts. The description of these contexts and their correlation with specific linguistic features presupposes the adoption of a view of the history of South Asian languages which does not operate in terms of a linear historical flow of homogeneous language varieties. Rather we must assume that at any given time a speaker has open to him a range of linguistic options (formal styles, contrasting lexical items and phonological units, etc.) in which the selection of particular features is correlated with any of a number of social variables. Linguistic competence is by nature complex, and any change in a language can result from a change in the social circumstances of language use as easily as from an internally motivated change in the code itself.

4.5. Conclusions and desiderata

In this chapter we have looked at the notion of "linguistic area" and discussed its applicability in South Asia. It is clear that there are a large number of linguistic traits which are shared by languages of the area without regard to their genetic stock, and that there are reasonable grounds for postulating linguistic transmission of these features across language boundaries. These features have been sufficiently pervasive in South Asia to warrant the suggestion that areal descriptions can serve as a useful supplement to classifications of the region based on purely historical-comparative criteria. We have also seen that there are numerous linguistic features which have been transmitted across language boundaries in South Asia, but which have not necessarily been adopted by the majority of languages in the area. This enables us to consider the usefulness of setting up a number of smaller-scale linguistic areas in South Asia.

This chapter has also focused on some of the history of the linguistic area, and noted that the transmission of linguistic features across genetic boundaries dates back to the earliest periods of Indo-Aryan residence in India. Historical evidence is sufficient to show that many South Asian areal features are not the result of the language-internal development of traits which would have occurred anyway. That is, multi-lingual contact was indeed the stimulus for the transmission of linguistic features across genetic boundaries--in some cases involving the adaptation of material already present in a language family, and in others involving the wholesale borrowing of such material.

The chapter also included a discussion of the extent to which the set of shared traits displayed by South Asia are unique to the area so that South Asia forms a typologically distinct area of the world. We have seen that according to Masica's analysis, South Asia does form a distinct area, set off by thick boundaries of isoglosses with regard to a number of word order phenomena, and has as its closest typological connection the Altaic language areas. The results of Masica's investigations, however, must be taken with great caution, as the typological markers he has plotted are for the most part quite different from the linguistic features which have been discussed in the literature on the South Asian linguistic area.

We also have looked briefly at the results of preliminary attempts to piece together the linguistic prehistory of the South Asian linguistic area. All such attempts have been hindered until recently by the lack of a sufficiently powerful model of sociolinguistic variation in which to frame the discussion of the transmission of features. We have seen that Southworth's research has opened up new avenues of approach for discussing linguistic areas as a reflection of social and linguistic stratification within a complex society. His approach will hopefully serve as a basis for more detailed investigation of the social conditions of linguistic borrowing in the history of South Asian languages.

In spite of the considerable progress which has been made since the influence of non-Aryan on Indo-Aryan languages was first pointed out, the areal study of South Asian languages clearly has many important problems before it. Our inventory of linguistic features which may have areal status is by no means complete. Data is missing for many languages, particularly Munda and Tibeto-Burman; and we simply do not have the full facts concerning the distribution of those areal features which have been identified. There are few dialect maps with isoglosses marking the dissemination of many of these linguistic features. Even in cases in which a feature has been identified in a given language, there is little information in regard to whether it exists in all strata of usage, or if is is restricted to particular social contexts.

Much more information about the histories of many of these features is necessary, so that more reasonable hypotheses concerning their origins and paths of transmission can be constructed. This is turn requires far more information concerning the sociology of language use in ancient times. History, unfortunately, reduces complex facts to apparent simplicity. The history of South Asian languages is essentially a history of standardized written records. At any period such levels of language use will of course be integrated into a complex network of linguistic styles and levels; but time has eradicated most of our knowledge of this complex network, and left us with deceptive impressions of language use. Clearly some sociolinguistic reconstruction is in order to fill in what Southworth calls the "linguistic stratigraphy" of the region.

Also of fundamental importance for the future is some sort of standardization of terminology and purpose. All South Asian arealists are looking for shared linguistic features, but there are few attempts to state what is meant by a linguistic feature. As should be all too clear from Chapter 2, the notion of "linguistic variable" is inherently a function of fundamental views on the nature of language and how best to go about linguistic investigation. Linguists from different schools and eras disagree as to such fundamental notions as what constitutes a linguistic "fact," and what types of "facts" should be compared for typological and classificatory purposes. This lack of agreement has all too often been reflected in work on South Asian languages. The establishment of a common set of ground rules for comparison of areal traits is highly desirable.

NOTES: CHAPTER 4

1. For a general discussion of linguistic areas, see Becker, 1948; Jakobson, 1931; and Bonfante and Sebeok, 1944.

2. Cf. also the following definition from Becker: "Unter einem Sprachbund verstehen wir ein Gruppe von Sprachen, die durch gemeinsame Schicksale im gleichen Kulturraum und durch wechselseitige Beeinflussung einander so stark angenähert wurden, dass man in jeder von ihnen ungefähr das gleiche auf ungefähr die gleiche Art sagen kann." (Becker, 1948: Vorwart, no page).

3. We can point to the history of the description of retroflex consonants in South Asian languages by way of example. August Friedrich Pott pointed out the widespread use of these sounds in the non-Indo-Aryan languages of the region as early as 1833, and went so far as to suggest that the sounds were acquired by Sanskrit from these languages. The Dravidian origin of Sanskrit retroflexes were also pointed out by Caldwell, Morris, Benfey, and Ascoli in the middle of the nineteenth century. For a thorough review of the history of the description of retroflexion in South Asian languages, see Kuiper, 1967:136-8. The discussions of other shared linguistic traits is of similar antiquity.

4. Eg. H. *pānī-vānī* 'water or something like it' [data supplied by MCS and HFS]; B. *ghoḍā-toḍā*, Mai. *ghorā-torā*, H. *ghoḍā-uḍā*, G. *ghoḍo-boḍo*, M. *ghoḍā-biḍā*, Si. *aśvaya-baśvaya*, Ta. *kudirai-gidirāi*, Ka. *kudure-gidure*, and Te. *gurramu-girramu* 'horses and the like' (data from Chatterji, 1953:49).

5. For further discussion of the so-called "echo words" in South Asian languages cf. Emeneau, 1938 and Ramamurti, 1931.

6. Cf. Emeneau, 1944.

7. In Kota the onomatopoetic forms have a basic CVC phonological shape

which can occur alone, with any of a limited number of derivative suffixes, or reduplicated (both with and without derivative suffixes). Where reduplication of the basic syllable occurs, it may be total, or it may be in conjunction with a change of vowel or with the initial consonant of the syllable, or both.

Examples:
A. Non-reduplicated.
1. CVC with no derivative suffix: *paṭ* 'Noise of bursting of skin when burnt, of sharp blow, crack' (3178).

2. With derivative suffixes: CVC - *k*, *burk* 'noise of flying up from ground' (3552); CVC - *a·r*, *caṭa·r* 'noise like whip-crack' (1893).

B. Reduplicated.
1. Identical reduplication with no derivative suffix: *cik cik* 'noise of a bird chirping'; *cur cur* 'noise made by meat when roasting' (2237).

2. Identical reduplication with derivative suffix: CVC*k*-CVC*k* and CVC-CVC*k*, *cork cork in-* 'to make noise in walking over fallen leaves or in bushes' (1946); *kad kadk in-'* 'heart, mind' beats fast with guilt or worry'.

3. Reduplication with change of vowel: *car cur in-* 'to make noise of a snake's motion' (1946).

4. Reduplication with change of vowel with derivative suffix: CV$_1$Ca·r-CV$_2$Ca·r, *doba·r daba·r* 'Noise of thrashing about while struggling' (2496).

5. Reduplication with change of initial consonant C$_1$VC-C$_2$VC, *ve·k me·k in* 'to make sport of someone'.

6. Change of vowel and of initial consonant with no derivative suffix; *jaṭ boṭ* 'noise of sexual intercourse'.

7. Change of vowel and of initial consonant with derivative suffix: *caṭr poṭr* 'noise like a whipcrack or loud crackling of fire' (1893).

8. Cf. Emeneau, 1944:[illegible]5-29; Hoffmann, 1952, Caldwell, 1856:554.

9. For an excellent discussion of the difficulties involved in trying to test the psychological validity of sound symbolism see Brown, 1958.

10. Examples of these four points (from Bright) are:

(1) LTa. *kir̤avi* 'old woman', CTa. *keḷavi; *ed-ir* 'opposite' >

OKa. *id-ir, ed-ir;* LTa. *uṭaṉē* 'immediately', CTa. *oḍane;* Tu. *kirè* ~ *kerè* 'tank', *munè* ~ *monè* 'point'; To. *er-* 'scatter', Ta. *iṟai,* Kod. *kere* 'to restrain; a tank', Ta. *ciṟai.*

(2) Vellore Ta. *po·hũ* 'it will go', *pɔ·rã* 'he goes'.

(3) Early Dharwar Kannada had a series of vowel phonemes with allophonic variants determined by the following vowel, including /ō/ which became [o·] before a high vowel, and [ɔ·] before a non-high vowel. Because of a later shift of /e/ to /i/ in final position, [ɔ·] acquired phonemic status, e.g. *[ko·ṭi] 'crore' > DKa. *kōṭi,* and *[kɔ·ṭe] 'fort' > DKa. *kɔ̄ṭi.*

(4) Te. *[gu·du-lu] 'nests' > [gu·llu], but *[gʊ·ḍɑ-lu] 'baskets' > [gʊḷḷʊ]; Konkani [kæ ḷæ̃] 'banana', [kẽ·lĩ·] 'bananas'; B. √šun 'hear' > *šona* 'heard'; A. *kola* 'black' (Skt. *kōkila*). *kɔla* 'plantain' (Skt. *kadala*); Si. *meheli* < Skt. *mahilā* (with mutual assimilation of [a] + [i] to [ə] + [e]).

11. In the first place, the Oriya language is a discontinuity in the supposed linguistic area. Furthermore, at least three types of assimilation have to be distinguished. Most of the Dravidian varieties, from the prehistoric period down to the apparently recent developments in Telugu, have involved partial assimilation to the open quality of a following low vowel. But outside of Dravidian, only Konkani fits this limited pattern. In Munda, the pattern is enlarged by the addition of progressive assimilation; in Bengali and Assamese, it is still further enlarged in that assimilation to high vowels also occurs; and in Singhalese, assimilation of fronting, as well as complete (rather than partial) assimilation, make their appearance. (Bright, 1966:322).

12. The following sentences exemplify the use of these constructions in Hindi-Urdu:

1. mãĩ bhī jānā cāhatā hũ̄. 'I also want to go'.
2. merā dost‿aur bhī hai jo. . . 'I have another friend who. . .'
3. us kā yahā̃ na āne par bhī. . . 'In spite of his not coming here here. . .'
4. vah caubīs ghaṇṭõ kām kartā hai. 'He works all the time (lit. all 24 hours).'
5. koī us se milne āyā. 'Someone came to meet him.'

13. Examples from Emeneau (1974a:112) are given with reference numbers from DED.

	Toda language		Kota language		Badaga language	
	male	female	male	female	male	female
Toda	o·l (342)	toz,mox (2885)	ton (2885)	tody (2885)	todava (2885)	toduvati (2885)

	Toda language		Kota language		Badaga language	
	male	female	male	female	male	female
Badaga	ma.f (3956)	madty (DEDS 3798)	mayv (3956)	mayt (3956)	baḍuga (cf. 4267)	baḍugati (cf. 4267)
Kota	kwï.f (1468)	kwï.ty (1468)	ko.v (1468)	ko.ty (1468)	ko.ta (1468)	ko.ti (1468)
Kurumba	kurb (1530)	kur(b)č (1530)	kavaŗ	kavaṛc	kuruma (cf. 1530)	kurumati (cf. 1530)

14. Cf. Greenberg, 1966.

15. Many Indian languages have constructions in which two independent verbal stems are combined to form a "compound" verbal stem whose semantic properties are slightly different from those of either of the components. Usually the meaning of the entire compound is some variation of the meaning of one of the components. Thus in Hindi: *mārnā* 'to strike, hit', *mār ḍālnā* 'to kill'; *honā* 'to be', *ho jānā* 'to become'.

16. Constructions involving perception, internal states of some kind, physical sensations, etc. in which the person doing the perceiving, sensing, etc. is put in its oblique case form and marked with an accusative/dative postposition and in which the surface structure subject is that which is perceived, experienced. Thus rather than English 'I hope that. . .' we have the equivalent of 'to me hope is that. . .'. E.g., H. *mujhee āshā hai ki*. . . 'I hope that. . .'; Ma. *eniKKu raamanE aṟinnilla* 'I didn't know Raman' (to-me Raman-Acc. knew-not).

17. Whereas English and many other languages use an explicit verb 'to have' in sentences such as 'I have a sister', 'she has a bad cold', 'we have a lot of work', 'they have to go', many other languages do not. In Hindi, for instance, the first of these sentences is expressed by a genitive expression (e.g. *merī ek bahin hai* 'my one sister is'). The others are expressed by dative constructions (cf. fn. 16) using the copula (e.g. H. *hamẽ bahut kām hai* 'us-acc. much work is'.

18. Hindi 20-1/2; Telugu 20; Bengali, Malayalam 19; Mongolian 18-1/2; Uzbek, Sinhalese 18; Japanese 17; Korean, Turkish, Kashmiri 16; Burmese, Amharic 15-1/2; Santali, Georgian 15; Hungarian 13-1/2; *Tibetan 12-1/2; Tajik 11; Russian, *Pashto 10-1/2; Czech 9-1/2; Persian, Chinese, Rumanian 8; German 7-1/2; Greek 7; Spanish 6-1/2; Swedish 6; French 5-1/2; English, Arabic 5; Swahili 3; Thai, Javanese 2; Cambodian 1/2. (Items marked with * have at least one feature for which value could not be assigned.)

19. For the acoustic bases of these oppositions, see Jakobson, Fant, and Halle, 1951.

20. E.g. H. *laRke ne vah pustak kharīdī* 'the boy bought (f.) that book (f.)'; G. *chokrie pāṭh vā̃cyo.* 'the girl read (m.) the lesson (m.)'; P. *tobi ne kapRe toe* 'the washerman washed (m.p.) the clothes (m.p.)'.

21. For recent work on the subject see Suseendirarajah, 1973 and Karunatilake and Suseendirarajah, 1973. Older studies include Kuiper, 1962; Shanmugam Pillai, 1962; Suseendirarajah, 1967; and Zvelebil, 1966.

22. Cf. Bloch, 1934:328-31.

23. E.g. Pa. *ţolavə́l* 'to collect', *ṭoṭá* 'piece', *ḍubē-də́l* 'to drown, sink', *ḍāḍ* 'calm, comfort'. Cf. Shafeev, D. A., *A Short Grammatical Outline of Pashto.* [=*IJAL*, vol. 30, #3], 1964, p. 5.

24. Cf. Southworth, 1974.

25. Cf. Kuiper, 1967:136ff.

26. Cf. Kuiper, 1967:136-8.

27. Cf. Kuiper, 1967:146-50.

Chapter 5

Social Dialectology

5.0. Introduction

The previous two chapters have been an examination of particular facets of the dissemination of language varieties in South Asia. In Chapter 3 we attempted to describe the distribution of linguistic codes through space, and to relate the traditional groupings of these codes into sets whose members share parallel genetic origin, while Chapter 4 examined linguistic properties shared by codes regardless of their historical origin, and considered types of intergroup contact and methods of transmission which could have lead to the sharing of such traits. We noted that the sharing of linguistic traits by codes in close proximity is probably a function of the social contexts in which individuals "control" portions or all of more than one of the juxtaposed codes. The selection of a particular code--or perhaps more accurately the selection of a particular register or style--is a reflection of the operation of social constraints.

In this chapter we turn our attention to these constraints. In speaking of "social constraints" we are, in fact, referring to situations in which it is possible to show statistical correlations between the alternates of purely linguistic variables and independently motivated social variables. An infinite array of such correlations is theoretically possible, and the determination of those social variables which yield the highest predictability of linguistic variation is a matter for empirical investigation. Unfortunately, the literature on social dialectology of South Asia has made little attempt to discuss the productivity of social variables. An overly large portion of the literature which is discussed here has relied too heavily on caste as the major independent social variable with which linguistic variation can be correlated, an assumption which has only recently begun to be challenged in the literature.[1]

In spite of these limitations of data, we discuss the correlations that seem to be justified on the basis of past research, and point out fruitful directions for future study. This includes an examination of studies which treat caste, age, sex, education, residence, and other variables as determining factors in linguistic variation. We also examine diglossia as a wide-spread social phenomenon in South Asia, take another look at diffusion of linguistic traits within complex sociolinguistic contexts, and discuss the sociolinguistic problems involved in defining languages in relation to their various pseudonyms.

5.1. General problems of social dialectology in South Asia

It has been noted since at least the early part of this century[2] that there exist overlapping linguistic codes whose use in South Asian society can be correlated with caste or other social variables. This correlation has been aptly described by John Gumperz as follows:

> A characteristic feature of Indian society is the segmentation of populations into ethnically distinct, endogamous groups or castes. These groups are only in part territorially separate. In many instances, they coexist in what social scientists study as a single community. They hold similar religious beliefs and regularly exchange services. Yet, although they are in constant communication, they may speak distinct languages and dialects at home. Whereas, in other areas of the world, intergroup communication in time tends to obliterate language differences, in India such differences appear to be in large part maintained. (Gumperz, 1969b:598)

Although linguists as far back as Bloch have recognized the pivotal role played by social variables in determining types of linguistic variation, most have been at a loss in coming up with coherent theoretical frameworks for describing complex sociolinguistic phenomena. Those attempts which have been made were hampered by inadequate theories for dealing with the range of sociolinguistic variation, by a failure to appreciate how linguistic innovations and subjective attitudes towards linguistic variation are transmitted, and by inadequate descriptive theories of the chains of communication which link the members of speech communities.

In addition to the above points, Gumperz raises a question which has been of central concern in almost all studies of social dialects in South Asia: what are the mechanisms by which correlations between linguistic and social variables are maintained? Many early studies of socially determined variation in language assumed the truth of Bloomfield's dictum that differences in linguistic code are associated with different patterns of social interaction--that groups which exhibit the greatest linguistic differences are those most isolated from one another (Bloomfield, 1933:47). As Bean points out, "the Bloomfieldian view of dialect was developed primarily with regional variation in mind, where rivers, mountains, and political boundaries separate groups and promote the development of linguistic differences. His view of social dialects is that isoglosses correspond to places in the social structure where communication is infrequent. . . . This concept alone will not account for linguistic variation in South Asia" (Bean, 1974:287).

In addition to noting problems arising from accepting Bloomfield's view of sociolinguistic interaction, Gumperz also questions the propriety of outright adoption of models of social structure taken from the experience of non-South-Asian traditions.[3] We strongly feel that the patterns of class stratification evident in South Asia are sufficiently distinct from western patterns to prohibit the direct transfer of Western social variables to sociolinguistic analysis in the South Asian context.

Given the extent of linguistic variability in South Asia, it is not surprising that there are, over a period of time, changes in the codes used by different social groups. It is therefore interesting to study the mechanisms by which change in social dialects take place, whether these be changes in linguistic rules themselves, or in the social variables which influence the selection of one linguistic option over others. However, it is not sufficient to state, as do some, that a certain linguistic form or construction is more prestigious than others; we need to know both the reasoning and the process behind the transfer of these subjective evaluations of speech forms into objective patterns of linguistic variation.[4] Gumperz suggests that adequate solutions to the study of sociolinguistic variation will only be found by studying speech use in its social context: "It seems that the solution to the problem of linguistic diffusion may be found in more detailed empirical study of interactional norms which limit interpersonal contact. It is quite possible that caste has different effects on communicative boundaries in modern urban and in traditional rural environments" (Gumperz, 1969b:601).

The central concern of social dialectologists in South Asia has been to determine the most productive means of drawing linguistic isoglosses, and to select the social variables whose symbolic mapping is most nearly congruent with the patterns of linguistic variation. An additional concern is with subjective attitudes toward linguistic variation. Bean has correctly noted that differences which a linguist perceives between the codes of two social groups may not be noted or held significant by the speakers themselves (Bean, 1974:291). This distinction in perception has too often been ignored, and the literature is replete with observations by individuals who pass on their subjective evaluations of linguistic alternatives as though they were confirmed objective patterns of variation.

5.2. *Studies of caste dialects*

By far the most widely studied social differences in language use in South Asia are those which are determined by caste. It is clear that caste dialects in the northern portion of the subcontinent are considerably different from those in the South. In the North, the main differences between caste dialects are between *savarṇa* and *avarṇa* (i.e. touchable and untouchable) groups, while the major distinctions in the South are between Brahman, non-Brahman and Adi-Drāviḍa (untouchable) caste dialects. Social dialectological literature has been more sensitive to caste dialects in the South, and has tended to view caste as the major sociolinguistic variable. Because of the lesser degree of clarity of *savarṇa* and *avarṇa* distinctions in the north, students of north Indian social dialects have been better able to isolate variables other than caste. Bean sees the difference between north and south Indian caste dialects as involving the notions of "purity" and "pollution": "The difference between North and South lies in the position of southern Brahmans. There they are more exclusively the agents of purity being the only twice born varṇa; there they have had a more exclusive contact with foreign languages and written literatures; and there they live separated from the rest of society" (Bean, 1974:286).

The earliest, and for many decades the only, study on caste dialects in India is Bloch's "Castes et dialectes en Tamoul" (Bloch, 1910). He notes that dialect stratification has existed in South Asia at least since the time of early Sanskrit theater, where high caste men spoke Sanskrit but women and servants spoke one or another form of Prakrit.[5] Bloch points out that this social differentiation escaped the attention of early scholars working on South Indian languages.[6] He concludes that there exists a Brahman/non-Brahman/untouchable distinction in social dialects in the Tamil area and states that this differentiation is not surprising, given the conservative nature of South India.[7]

After cataloguing phonological, morphological, and lexical features which differentiate some of the castes,[8] Bloch concludes that (1) Brahmans are more phonologically conservative than other groups; (2) morphological variation is more diagnostic of language stratification than other types of variation because of its regularity; (3) and there is a definite stratification of language varieties in Tamil correlated with caste. While the upper and lower strata of this pattern are stable, the middle castes tend to be less so, primarily because of the attempts at upward social mobility in these groups.

It is interesting to note that although Bloch accepts the existence of linguistic variability among castes, he never claims that the differences among caste dialects are to be considered categorical. He considers Tamil dialects to form a graded scale with regard to their resistance to innovation: "Il y a. . .une échelle de résistance aux altérations phonétiques, suivant les castes; ainsi s'explique que la hiérarchie de correction des formes sont en gros la hiérarchie des castes elle-même" (Bloch, 1910:14).

Bloch is concerned with the subjective impressions of caste dialects among Tamil speakers. He asked his informants whether they could tell the caste of another person while blindfolded, and their reply was affirmative.[9] He also considers the possibility that social variables other than caste may play a role in the determination of Tamil social dialects, and intimates that education and life-style have an effect on the speech of middle castes; but he specifically denies that education might serve as a social leveler.[10] (It is possible that in 1910 education was not yet widely enough available to middle caste Tamilians to act as a major influence on the speech patterns of large numbers of speakers.) Bloch also notes that since women were not being educated in any appreciable numbers, whether in Sanskrit, Tamil, or English, they exerted a conservative influence on the social dialects of their time. Bloch's work thus sets the scene for much of the future study of social dialects, particularly in South India, but elsewhere in South Asia as well. His major claims--that caste and dialect are associated, that there are three main socially determined Tamil dialects, that the middle castes exhibit greater variation than low castes,[11] that caste dialects differ in their degree of resistence to phonetic alterations, and that education, sex, and life-style ("genre de vie") have some effect on dialect variation--have initiated the serious study of social dialects in South Asia.

Studies of caste dialects[12] are generally confined, as we have noted, to Dravidian languages, particularly Tamil, Kannada, and Tulu. McCormack (1960, 1968a) has studied Dharwar Kannada dialects paralleling Bloch's Tamil study and noted a three-way distinction among Brahmin, non-Brahmin, and Harijan dialects. McCormack played tape-recorded samples of the speech of different speakers to a sample of Kannada speakers. His subjects were asked to identify the caste of the speakers in the recordings. He found that his subjects could correctly identify the caste of the speakers in the tapes 40% of the time, while 35% of the responses were incorrect and 25% of the subjects failed to respond at all. Brahmin speech was most often correctly identified, while Harijan speech was almost never distinguished. McCormack concludes that Harijan speakers either mask their speech by using other forms, or that they do not use the stereotyped forms other speakers expect of them.[13]

In the latter study, McCormack (1968a) attempts to determine the cause of caste dialect differences, that is, to determine where they originate and the direction in which they spread. He attempts by statistical analysis to show that "variations from expectable language behaviour in the Brahmin group induces change in the speech of non-Brahmins." (McCormack 1968a:223) Furthermore, "the mechanism by which Brahmins induce linguistic deviation among non-Brahmins, and therewith also effect variability in the dialects of members of other castes. . . is most commonly through openly correcting the speech of their own children." He concludes that

> studies of caste dialects have proceeded so far, then, as to indicate that Indian speech communities are conscious of caste

> dialects, that linguistic forms diffuse more rapidly among lower-status groups, that the speech of lower-status groups displays forms with wider regional dispersion than does the speech of locally high-status castes, and, finally, that the motivation for the existence of caste dialects is provided by the social status aspects of the local caste hierarchy itself. (McCormack 1968a:225)

These are stong claims. That Indian speech communities are conscious of caste dialects seems self evident; but the claim that linguistic change occurs more rapidly among lower-status groups than among high-status castes requires much more justification than has been given. Different characterizations of the relative homogeneity of Tamil caste dialects have been given elsewhere. Ramanujan (1968) shows that Tamil Brahmin speech has fewer geographically determined variants than does non-Brahmin Tamil.[14] Ramanujan also demonstrates that phonological change is more characteristic of Brahmin dialects than of other dialects, and that non-Brahmin dialects tend to display innovation in other ways.[15] McCormack's claims about the amount of innovation in Kannada is not supported by other Dravidian data and further research is needed to show if his observations are restricted to Dharwar Kannada or if they have support elsewhere.

Additional data on Kannada caste dialects is to be found in Shankara Bhat's study of caste dialects in the Mysore District (Shankara Bhat, 1967-8). He finds that "we may set up three social varieties, those of the Brahmins, of the untouchable castes and of the rest (local), all of which intersect the three geographical varieties. . .". (Shankara Bhat, 1967-8: 68). He characterizes the differences between the dialects in terms of a series of changes operating on the norm represented by the standard language.[16] The Brahmin dialect, although encompassing a fair amount of *regionally* determined variants, nevertheless displays the smallest deviation from this abstract norm.[17] Here again we must ask if the Brahmin dialects are closer to the standard than non-Brahmin ones because the standard is historically derived from the Brahmin dialects. As Friedrich points out, the Brahmins have had a monopoly on inscription and other writing from antiquity, and it is not surprising that their dialect closely resembles the standard dialect; but there is no reason to assume that non-Brahmin dialects were produced by historical processes of derivation from the hypothetical standard.[18]

Tulu social dialects are examined in Ramaswami Aiyar, 1932. The work enumerates the major differences between Brahmin and non-Brahmin Tulu. According to Ramaswami Aiyar, the Brahmin dialect displays more lexical borrowings from Sanskrit than does the non-Brahmin dialect, although both agree in having loans from Prakrit, Pali, and, of course, Kannada. As Tulu is spoken in a small area--primarily South Kanara District, Karnataka State--social differences among dialects are more apparent than regional ones. Ramaswami Aiyar attributes the differences between Brahmin and non-Brahmin Tulu largely to the "cultural aloofness of the Brahmin community." (Ramaswami Aiyar, 1932:898) It is not clear to us, however, how such hypothetical "aloofness" can be made to correlate with degrees of innovation in linguistic systems. Since there are no written records for Tulu which attest to early stages in the development of Tulu, it is an interesting problem for historical linguists to show which of the two dialects has diverged more radically from a reconstructed Proto-Tulu. We might add that tentative comparisons of Tulu caste dialect data seem to indicate that non-Brahmin Tulu has undergone greater deviation from Proto-South-Dravidian than have other dialects.[19]

Tamil social dialects, after remaining uninvestigated for a number of decades subsequent to Bloch, 1910, have been extensively studied in recent years. The most important studies on this topic are Andronov, 1962, Shanmugam Pillai, 1965a, 1968, Ramanujan, 1968, and Yesudhason, 1975. Andronov's monograph on spoken Tamil and its dialects is primarily concerned with geographical dialects of the language, but does contain some useful information on social dialects. His major concern is in contrasting Brahmin dialects (considered to be spoken by an urban intelligentsia) with middle and lower caste ones. It is his contention that Brahmin dialects show greater uniformity than do non-Brahmin dialects. He supports his claims with pronominal data, as Tamil pronouns display significant amounts of variation across castes.[20] As an overview of the entire dialect situation in Tamil, Andronov, 1962 is unexcelled.

Shanmugam Pillai has written several dialectological studies concentrating on lower caste Tamil dialects, dialects which are extremely difficult to obtain reliable data about. In Shanmugam Pillai 1965a, he provides a statistical index of caste isolation as a function of the use of particular variants of kinship terms. Except for the Tamil of Muslim groups, the greatest number of non-shared kinship terms is in the Brahmin dialect, with groups descending in caste ranking having increasingly less deviation from the shared core of terms. Shanmugam Pillai offers two possible explanations for these statistics. First, he suggests that the general lack of communicative interaction between Brahmins and lower castes obviates the need for shared terminology. Second, he asserts that Brahmins retain distinctive caste-marked terms because of the prestige involved. While higher non-Brahmin castes may seek additional prestige by employing Brahmin forms, lower ranking castes would receive nothing but ridicule for such emulation. Rather than adding prestigious caste-marked terminology, the lower castes prefer to rid their speech of caste markers which identify their low social status, resulting in a reduction of terminological differences in the lowest level social dialects.

In another study, one of a dialect of Kanyakumari fisherman, Shanmugam Pillai (1968b) employs statistical methods to determine whether the dialect studied bears its closest resemblance to Brahmin, "higher" non-Brahmin, or "lower" non-Brahmin dialects. The conclusion drawn is that the speech of Kanyakumari fisherman "cannot be said to be closer to any one particular dialect" (Shanmugam Pillai 1968b:726). This paper is of interest because, in addition to discussing the linguistic features which can be cited in establishing a Kanyakumari fisherman dialect of Tamil, and examining the relationship of this dialect to other Tamil speech varieties, Shanmugam Pillai also notes the existence of code switching among various dialects by Kanyakumari fishermen.[21]

The "motivation" for this code switching is seen to lie in "the hierarchy of the caste structure coupled with its prestige and politics and not any economic and educational factors" (Shanmugam Pillai, 1968b: 726). Unfortunately, Shanmugam Pillai fails to offer concrete proof of how "the hierarchy of the caste structure" is correlated in a direct way with alternations in dialect. The following explanation cited for code switching is unconvincing, although not on the surface implausible, and needs to be supported with more data than is given:

> The economically and educationally backward lower castes and the economically and educationally forward higher caste--the Brahmins, both switch on to a Higher Non-Brahmin dialect. . . . The anti-Brahmin movement resulted in the loss of prestige for the Brahmin dialect and even the Brahmins found it wiser to switch on to the

> Higher Non-Brahmin dialect although it does not elevate them in the caste hierarchy. (Shanmugam Pillai, 1967:726).

If "the hierarchy of the caste structure" motivates dialect switching, why do Brahmins, the highest group in the traditional hierarchy, switch codes? What need would they, presumably at the top of the social hierarchy, have to make their speech conform to different norms in certain social contexts? Unfortunately, the anecdotal format of this article makes it impossible to answer this question and to test hypotheses about correlations between linguistic variation and specific social variables.

Another study of Tamil caste dialects is Yesudhason (1975), a brief but interesting study of age-based variations in caste dialects of a regional variety of Tamil (the Vilavenkotu dialect, Kanyakumari) which is spoken near the Tamil-Malayalam language border. Some castes, particularly the Krishnavagai caste (highest) originally exchanged brides with some Malayalam speaking castes, and the Tamil of these groups shows features attributable to Malayalam influence. Younger speakers do not show these Tamil and Malayalam archaisms, but rather show features that are particularly local. The intermediate age group, interestingly, demonstrates greater congruence with the regional standard than does the youngest group; but among the older Harijans, there are few archaisms and more innovations. Here we see that age, caste, and marriage patterns combine to influence differentially the retention or innovation of features in Vilavenkotu Tamil. These data are supported by some unpublished work by Schiffman (1975) on Coimbatore non-Brahmin Tamil, where among certain middle to lower castes, older speakers show more uniformity across caste lines than within. Again, their forms are conservative, showing congruence with Literary Tamil, and in some cases, with Brahmin dialect.

Yet another study of Tamil caste dialects is Ramanujan (1968), in which are compared Brahmin Iyengar and non-Brahmin Mudaliyar Tamil with historical innovation noted in both dialects. The two dialects differ in the nature and scope of this innovation. Brahmin dialect generally innovates in ways which create lexical and phonological contrasts, while the non-Brahmin dialect innovates by leveling paradigms, thereby causing its speakers to be more homogeneous than the speakers of literary standard Tamil. A problem in this sort of analysis is that statements of "innovation" in competing dialects presuppose knowledge of the ancestral speech varieties from which the contrasting dialects are derived. More than likely, the early literary variety of Tamil used as a basis for comparison was itself based on the then current Brahmin dialect; however, lack of certainty in this respect opens the possibility of inaccuracy.

It has been argued that Brahmin morphological conservatism is a function of literacy. In an article comparing Tamil and Tulu, Bright and Ramanujan (1962) have attempted to determine the importance of literacy in language change in caste dialects. They conclude that caste dialects innovate independently of each other, with "conscious" change (semantic shift, lexical and phonological borrowing) being prominent in the upper class dialects, and "unconscious" change being typical in the lower caste dialects. "In some. . .examples, B and NB seem to have innovated equally, but in different directions. . . . Neither dialect has a monopoly on innovations. . .yet tendencies are discernible: on the part of B, toward greater use of foreign vocabulary, foreign phonological and semantic shifts; on the part of NB, toward shifts in native phonology and morphology" (Bright and Ramanujan, 1962:1111).

In Tulu, the non-Brahmin dialect has greater phonological innovation than the Brahmin one, although Brahmin does have a rule which aspirates consonants. In morphology, both Brahmin and non-Brahmin are found to innovate.

> In summary, the Tulu evidence shows the Brahmins as chief innovators in the more conscious varieties of change--semantic shift, lexical borrowing, and phonological borrowing. In the less conscious processes of phonological and morphological change involving native materials, both B and NB dialects innovate.
>
> We feel that the evidence so far examined supports the hypothesis that upper and lower class dialects innovate independently of one another, and in two ways, here labelled conscious and unconscious. Of these types of change, the more conscious variety is regularly the mark of the upper class dialect. . . . in Kanarese and Tamil, where there is widespread literacy among Brahmins, the formal written style seems to have retarded the less conscious processes of innovation . . . We feel that further investigation of social dialects in the South Asian context can contribute much to understanding the mechanisms of linguistic change. (Bright and Ramanujan, 1963:1112).

Bright and Ramanujan are thus pointing out that linguistic change in the South Asian context is illustrative of some general processes, i.e., there are social factors which are clearly involved in the South Asian scene that may also be important in other parts of the world.

In other words, theories of language change which ignore these factors are probably lacking in generality. We should add to this that insights about the mechanism of linguistic change such as those provided by Labov (1972:178-80) also need to be applied to South Asia so that the results of their application might contribute to general sociolinguistic theory.

In summarizing the discussion of South Asian caste dialects in past literature, we see that scholars have generally been interested in how independently observed distinctions such as those of caste are associated with formal linguistic differences; and secondly, how and why these differences are maintained. Since the Bloomfieldian notion of density of communication as *the* determining factor in dialect innovation or conservatism is obviously not fully applicable to South Asia--since differences in South Asia are maintained despite dense communication, sometimes over centuries of close contact--scholars have had to develop other theoretical frameworks to explain caste dialects and their rates of change in the subcontinent. The failure of the Bloomfieldian model to predict linguistic variation as a function of the density of communication between individuals has been well summarized by John Gumperz as follows:

> One of the most important results of recent studies in speech variation so far has been the clarification of the relationship between intensity of communication and the assimilation of linguistic forms. Bloomfield's assumption that intensity of communication leads to a decrease in speech variation is only partly justified. In highly stratified societies such as the caste societies of India, it is quite possible for people to be in constant and regular communication over long periods of time without adopting each other's speech patterns. It would seem that communication leads to uniformity only when there is both the possibility and the desire for social assimilation. Where social norms put a premium on social distinctness, linguistic symbols of such distinctness tend to be maintained. (Gumperz, 1967:227-8).

If the Bloomfieldian hypothesis concerning the basis of linguistic variation has proved inadequate for explaining caste-based variation in South Asia, new models have not appeared to replace it. Caste studies of South Asian dialects have operated in somewhat of a theoretical void, with little conscious effort made to justify the theoretical assumptions underlying the establishment of dialects. Among most promising applications of general sociolinguistic theory to South Asia might be the application of the sort of variability theory being advanced by Labov, Bickerton, DeCamp and others,[22] but until quite recently, those proficient in contemporary variability theory seldom have had either sufficient knowledge of South Asian languages or access to speakers of the dialects necessary for such research. Moreover, the application of new models of variability theory is valuable precisely because it requires an examination and justification of the social variables chosen to serve as linguistic differentiators. When variability theory is made explicit, it should always be possible to ask whether a given social variable is the most productive one possible, or whether it might be observing more fundamental social conditioning factors. This sort of examination has not occurred in the case of South Asia, and the predominance of caste studies over studies of other social indicators may be not so much a reflection of the importance of caste as a social indicator, but a result of a failure to systematically look at other possibilities.

5.3. *Social variables other than caste*

Although caste has been the most frequently described social variable which is correlated with observable linguistic differences in South Asia, it is clear that there are other social factors--education, sex, age, residence (urban vs. rural), etc.--which are strongly associated with patterns of linguistic diversity. It would seem that the overriding concern of social scientists with caste in South Asia has rendered difficult the recording and description of other social variables and given an unbalanced picture to the total set of social constraints on language use. This has, as mentioned earlier, been more true of the literature on South India than of other areas, but the thrust of these points is true for all of South Asia.

In this section we turn out attention to the limited literature available concerning sociolinguistic variables other than caste in South Asia, and recapitulate criticisms made by recent scholars on the consequences of assuming an overly caste-conscious viewpoint. Our aim here is not to downgrade those studies which have examined caste, but to stimulate interest in the study of an array of social variables.

In a 1962 paper, M. L. Apte attempts to trace the development of a standard spoken Marathi and establish which subjective factors are most important to Marathi speakers in establishing a ranking of dialects according to prestige. Two factors, perceived urbanness and perceived education seem to be the most important criteria used by Marathi speakers in evaluating the speech of other Marathis. Marathi speaking informants, when asked to rate the educational and residential backgrounds of tape-recorded speakers, were easily able to establish whether the subjects were urban, rural, or from the Konkan. Brahmin caste was identifiably distinguished from non-Brahmin, and the educational level of the speaker was likewise readily surmised. Phonological differences between taped subjects seemed to be the primary basis of the informants' judgments. This contrasts with a study by McCormack (1960), in which the author shows that Kannada informants identified the caste of speakers on the basis of morphological differences. We require further investigation to determine whether the differences between the Marathi and Kannada cases are peculiar to the languages involved, are related to the specific social variables investigated, or are due to errors in observation and analysis.

One study which does make an attempt to determine the social parameters of linguistic variation in a community is Gumperz' "Dialect Differences and Social Stratification in a North Indian Village" (Gumperz, 1958). Rejecting Bloomfield's idea that the linguistic diversity in a group of individuals is inversely proportional to the degree of social contact and interaction between those individuals, Gumperz attempts to find out whether residential patterns, the ritual purity of individuals and groups of individuals, occupation, adult friendship contacts, children's play groups and, of course, caste are significant indicators of sub-dialects in a north Indian village, Khalapur, located in the Saharanpur District of Uttar Pradesh and having a population of about 5000. He concludes that "there is some correlation between the linguistic groupings and ritual status. . . . By examining inter-caste communication, we find that linguistic differences have no correlation with work contacts. . . . In the present study, the determining factor seems to be informal friendship contacts." (Gumperz, 1958:44) Gumperz is able to point to the existence of six distinct speech varieties in Khalapur and to associate these varieties with isolatable social groups. These social groups are not enumerable in terms of single social variables, and involve the interaction of occupation, age, religion, caste, and area of residence. The six groups are as follows:

(1) Hindu and Muslim touchable castes, except for "old-fashioned"

individuals and Rajput residents of two specific areas of the village;

(2) Rajput residents of two particular areas of the village;
(3) "Old-fashioned" individuals of all touchable castes;
(4) Chamars [a group of landless laborers];
(5) Shoemakers;
(6) Sweepers.

The important point to be observed in this study is that Khalapur, a rather small area of social interaction, demonstrates a complex structure of linguistic variation, and that in order to offer explanations for this structure the settlement history of the village, the demographic properties of individuals, and patterns of individual and group interaction need to be considered.

In a 1968 study (McCormack, 1968b), McCormack attempts to correlate occupation and residence with linguistic variables in Dharwar Kannada. He sought to determine whether the speech habits of Brahmins are influenced by prolonged residence in predominantly non-Brahmin areas and vice versa. He concludes that "no statistically meaningful relationships emerge between Brahmin residence in non-Brahmin residential areas and the frequency of Brahmin adoption of non-Brahmin traits" (McCormack, 1968b:480). McCormack was unable to find evidence indicating a willingness of non-Brahmin Kannada speakers to modify their speech patterns on the basis of residence. Rather, he found a correlation between occupation, particularly a white-collar occupation, and the adoption of Brahmin linguistic traits by non-Brahmins.

During the past few years articles have appeared which have advocated the systematic study of sociolinguistic variation motivated by social factors other than caste. Pandit (1967) has correctly observed that the past preoccupation with noting caste-based linguistic differences presupposed an overly static view of modern South Asian society. He advocates the study of "developing urban centres [which] display emerging social classes with considerable vertical mobility. . ." (Pandit, 1967:218). Unfortunately the Pandit article is confined to a critique of the past orientation of sociolinguistic studies of South Asian languages, and offers few demonstrations of the operation of a wide range of social variables.

A recent article by Pattanayak (Pattanayak, 1975) takes up the theme raised by Pandit and criticizes the failure of scholars to look beyond caste in the determination of social dialects of Dravidian languages. Pattanayak catalogues the linguistic features which had been previously cited as differentiating among caste dialects, and demonstrates that many of these are not categorically used by members of the castes with whom they are supposed to be associated. Like Pandit, Pattanayak correctly observes that urban locales are optimal areas for the study of sociolinguistic variation. He advocates the study of linguistic variability in which we assume dialects to be in states of flux, and in which numerous social variables exert varying influences on the utilization of various alternatives of linguistic variables. He notes that "from the linguistic point of view it is more important and interesting to study linguistic variation in differing social contexts rather than making static categorical statements about linguistic differences" (Pattanayak; 1975:102).

It is clear from the above discussion that although the desirability of incorporating a wide variety of social variables into sociolinguistic studies has been pointed out by investigators, the actual carrying out of such labor has hardly begun. This is not unnatural seen historically. The study of the "dialects" of various South Asian languages has had its theoretical underpinnings in Western dialectology. It has been recognized for hundreds of years that the major South Asian languages have regional varieties. The systematic examinations of social criteria of dialect formation, however, is of recent origin, and it is hardly strange that the transfer of these investigations to South Asia has concentrated on caste, the most obvious social variable in South Asia. But the dialects produced by adopting caste as a sociolinguistic parameter have in the past been rather "static" areas, with the tacit assumption that once the caste of an individual is identified there is a set of linguistic features which the person will employ. It is almost as if the Stammbaum characterization of the history of language families were expanded to include binary bifurcation of modes into caste dialects.

It seems to us then, that we have reached the point when investigators ought to begin to construct dynamic models for sociolinguistic investigation in South Asia. It is not difficult to see how this might be carried out. First of all, it is imperative that scholars begin to operate within frameworks which assume that variability in linguistic structures is a systematic feature of language. Rather than assume that caste is a social property which allows the postulation of invariable dialects--where a person either speaks the dialect or does not--we can start to consider it as one of a large class of social parameters which may or may not be statistically correlated with specific types of linguistic variability in particular contexts.

Once one adopts such a framework it becomes a matter of empirical investigation to determine which social variables are likely to have the greatest predictive force for linguistic variation. But such variables are highly unlikely to have categorical effects in all contexts. They surely will interact with one another to produce complex effects, and the force of single variables is likely to be revealed by sophisticated types of factor analysis. Moreover, it is an open question about the extent to which the categories that will ultimately emerge in such studies will resemble the categories that have proven useful in sociolinguistic investigations of North American urban centers.

The above discussion in this section has made it obvious that these questions are the most serious ones facing the next decade of research in South Asian sociolinguistics. We have been able to shed here only the dimmest light on the sort of variables which may eventually be of significance. It is likely that several other variables--sex, age, generation, education--will emerge as significant barometers of linguistic variation, depending on the outcome of future studies.

5.4. *Diglossia*

There is another type of linguistic variation in South Asia that has received a fair amount of attention in the literature, and that involves a split standard of usage by a community of speakers, where the extremes of this split are each associated with a definable set of contexts. This divided norm, referred to as "diglossia", has traditionally been defined as the phenomenon whereby languages exhibit two or more distinct styles of speech, one, sometimes called L(ow), informal or colloquial, is used by people for everyday use in their homes, in the marketplace, for making jokes, and in a large range of informal contexts. The other, called H(igh), formal or literary, is used in public speeches, prayer and other religious or ritual uses, and is the style that is the vehicle of literary traditions and is commonly taught in educational institutions. Diglossia as a field of inquiry for South Asian sociolinguistics was established by Ferguson in an article (Ferguson, 1959) that both invented the term and laid the groundwork for the study of the phenomena indicated by it. There have been a number of studies of diglossia in those South Asian languages that exhibit it to a significant degree, where it is often the case that mutual intelligibility between the H form and the L form(s) is restricted. That is, illiterates or semi-literates of a severely diglossic language like Tamil will experience difficulty in understanding many varieties of H Tamil. Although the phenomenon has been widely reported in Tamil, it exists to a greater or lesser extent in many South Asian languages. It even seems to be the case that for languages which historically have lacked it, e.g. khaṛī bolī Hindi, an H form (Sanskritized Hindi) has been invented to fill the gap.

M. Shanmugam Pillai has written a number of studies of diglossia in Tamil, mainly of a descriptive nature. In one (Shanmugam Pillai, 1960) he compares two norms, the variety of literary Tamil that he ordinarily writes and speaks (in formal contexts) with the variety of spoken Tamil that he uses in informal situations. He notes that within the norm called Literary, there are differences of style--"rhetorical" and "Pandit" Tamil are examples of these. Of course in spoken Tamil there are also regional and social varieties that would only be used in informal contexts. Thus there is a polarity between two kinds of Tamil, defined only as those varieties used formally and those varieties used informally, even though some of the formal varieties might never be used by certain speakers, indeed they might not even understand some of them, especially Pandit style. Similarly, the spoken varieties would not all use most of the regional and social dialects that exist. But for all speakers the polarity exists, and Shanmugam Pillai has taken the two varieties he controls and systematically outlined the phonological and morphological differences between them.

Since the linguistic theory within which this study has been done is American structuralism, we naturally have certain correspondences between the two forms handled in ways that are typical of that theory. For example, he describes the absence of a given phoneme in a given context as seen in Figure 55. Since American structuralism regularly deals with surface phonetics only, the generalization that /y/ is deleted after long vowels in final position, but is present elsewhere

CV:y regularly corresponds to CV

Literary Tamil	*Colloquial Tamil*	
pa:y	pa:	mattress
te:y[28]	t̲e:	rub
ka:y	ka:	nut
na:y	na:	dog
no:y	no:	disease
va:y	va:	mouth

Figure 55. Phonological Correspondences in Literary and Colloquial Tamil. (Shanmugam Pillai, 1960:29)

(e.g. if a suffix such as the clitic /um/ "and" were to be added to the above spoken forms, the /y/ would reappear) is not captured. In other words, a surface-phonemic comparison of the two varieties of Tamil is unable to show that in many cases, the underlying forms of the spoken forms are closer to the literary than is apparent from the surface phonetics. In another example, he shows that the literary form /paṭam/ "picture" has the spoken form /paṭon/ with the final /n/ representing nasalization. In actuality, the final /m/ of the underlying form of this and other similar items conditions the rounding of /a/ to /o/, before nasalization of the vowel to [õ]. Because of his reliance on Bloomfieldian theory, Shanmugam Pillai's presentation of the differences between literary and spoken Tamil obscures the difference between phonological and phonetic processes, and also ignores the optionality of certain rules, or the contextual variation in certain phonological environments.

Similarly, Shanmugam Pillai's morphological analysis is mainly a tabulation of allomorphic differences within single morphemes. He notes, for example, that "the plural suffixes are different whether they occur finally or non-finally." (1960:37) He compares L(iterary) T(amil) and C(olloquial) T(amil) plural endings and shows that the LT endings are the same whether they are final or non-final, whereas the CT endings are not, as noted above. However, it is obvious that the non-final CT endings are the true underlying forms of those morphemes, and the form which occurs in final position has been operated upon by a number of general rules, e.g. the deletion of final /ḷ/ and the rounding of /a/

before /ḷ/ before its deletion.[23] Primarily Shanmugam Pillai notes that by his analysis, LT generally has fewer allomorphs than does CT, due, of course, to certain phonological processes that are part of the grammar of CT that are not found in the grammar of LT. He also claims that the colloquial form and the literary form "seem to represent two different stages of development of the language existing side by side." (Shanmugam Pillai, 1960:40) This is a claim which we find requires much more substantiation than has been offered so far. (cf. Friedrich, 1961; Schiffman, 1970b) It is true that LT and CT are probably closer in underlying forms than Shanmugam Pillai's analysis seems to indicate, yet in other ways they are very different, especially in terms of late phonological rules that are part of the grammar of CT.

Shanmugam Pillai's 1960 article, in spite of our criticisms just given, does raise some interesting points about the almost absolute uniformity of Literary Tamil. One of the most important of these is that a speaker cannot be identified as to caste or region by the variety of Literary Tamil he uses, a fact which is not true for his colloquial speech. Shanmugam Pillai also points out the fact that attaining literacy in a highly diglossic language presents more problems than languages without such stylistic cleavage.

In a later article (1965b) the same author makes the claim that colloquial Tamil is beginning to displace LT in some contexts, e.g. in the film, in the dialogues of novels, and occasionally in public speeches. For instance, the beginning and ending of a speech are always in literary style, but in the middle of a speech, especially for the purposes of joking or punning, the colloquial style may be used.[24] This seems to us to be not a merger of the two styles, but an enlargement of the number of contexts in which colloquial may be used, and a diminution of the contexts for LT. Merger of the two styles, at least in our understanding of this term, would mean the development of a norm which is intermediate between the two, e.g. a plural form /-koḷ/ instead of LT /-kaḷ/ or CT /ko/. Since such forms do not occur, and should properly be starred, merger of LT and CT is not an apt description of Tamil diglossia.

In another article that treats, among other topics, Tamil diglossia, Zvelebil (1964) delineates the different varieties of speech exhibited in Tamil Nadu, including caste and geographical dialects as well as diglossic varieties. Zvelebil feels that the ternary distinction in caste dialects found by other scholars is wrong, and "would rather say that the binary division Brahmin vs. non-Brahmin is fundamental and basic in Tamil." (1964:240) He says that there are features common to the language of uneducated people (*koccai* or vulgar Tamil), not correlated with caste hierarchy but merely with education and profession. He would thus relegate the third and lowest category of caste dialects to a kind of style associated with disadvantaged speakers of whatever caste. This is a claim that cannot be dismissed out of hand, although it requires more substantiation.

In the sections of this article devoted to diglossia, Zvelebil gives an excellent description of various facets of Tamil diglossia, as well as some theoretical discussion of how to characterize it. He notes that colloquial Tamil has entered some domains previously restricted to it, such as on the stage, in the movies, and in broadcasting. He feels that "there is no doubt that spoken Tamil will, in a modified form, much like the Bengali *calit bhāṣā*, enter the field of creative and later even technical writing; and for a period of time, there will probably be two standards. . . . Gradually, however, the extremes may be moving more and more toward the *centre* which lies somewhere in the future as the one national language of Tamilnad." (Zvelebil, 1964:258-9)

Zvelebil's discussion of diglossia focuses on the common colloquial, differing from local and social dialects. He shows, for example, that the plural of the LT pronoun /avarkaḷ/ "they" has a CT form /avaŋka/ which is not found in the regional dialects:

> in South Eastern dialects the prevailing form is /avuka/, in South Western /avarkoḷ/, in Western /aviya/, in Ceylon /avaŋkaḷ/, in Brahmin speech /ava:/, in koccai /avuŋka/, whereas in the colloquial used by educated middle-class city-dwellers, it is /avaŋka/ which is nearest to the Eastern and Northern forms. (Zvelebil, 1964:259).

Zvelebil's main point is that CT is neither socially nor regionally based, but is a class rather than caste dialect evolving among middle class speakers of primarily non-Brahmin caste who are involved in mercantile and professional activities in urban areas.

While we agree that such a standard seems to be evolving, we do not share his optimism that this norm will supplant or replace Literary Tamil. A consequence of Zvelebil's position is that Sri Lanka dialects of Tamil, now linked with mainland Tamil only through the common acceptance of the 13th century literary norm, would necessarily be excluded from participation in the standardization of the new literary norm. This might lead to an independent evolution of a Sri Lanka literary Tamil, i.e., the birth of a new South Asian literary language. It is not unreasonable that both Indian and Sri Lankan Tamils, in an attempt to preserve the unity of mainland and Sri Lankan Tamil, might advocate the maintenance of the current literary language.

Finally, for Tamil, there exists a study by Schiffman (1973a) that focusses both on diglossia and on language and politics and which attempts to delineate the socio-cultural reasons for the persistence of Tamil diglossia. Tamil seems to be more resistant than any other South Asian language to any movement toward a colloquial norm. Schiffman asserts that Tamil diglossia is rooted in a kind of purism, which is related to the strength of the purity-pollution complex in South India. He also feels that the antiquity of Tamil literature, so important a value in South Asian culture, lends added strength to the purity myth, and buttresses the notion that Tamil is some kind of bulwark of archaic, unsullied Proto-Indian (or at least Proto-Dravidian) culture. Since Tamil has never borrowed aspiration in consonants, as the other Dravidian literary languages have, nor has it developed a contrast between voiced and voiceless consonants, there seems to be built-in resistance to borrowing from outside the language. This, coupled with the general knowledge that Tamils have of the antiquity of their literature and culture, leads them to believe that Literary Tamil has resisted change, and that a movement to dilute it with colloquial forms would be to rob it of its pristine purity and open the gates to a flood of Hindi and English influences. If what he claims is true, this would help to explain why diglossia is such an important phenomenon in South Asia, since some of the values inherent in the Tamil situation must be shared with other linguistic groups to a lesser extent.

Bengali is an example of a different trend, one of a gradual "merger" of the literary and colloquial styles. Dimock (1960) has delineated the history of diglossia in that language and the rise of a new literary norm converging from the older literary norm and the Calcutta standard colloquial. Interesting here is the influence of western forces, specifically the founding of Fort William College in

Calcutta and the attempts of its staff to train Bengalis in a form of the language that would meet the new needs of the language and especially of the East India Company. The influence of English prose style on Bengali and the disdain for Sanskritized Bengali of an earlier period is also mentioned, as is the manner in which Tagore and other writers made conscious choices that could not be ignored in the development of modern Bengali. Dimock outlines the kinds of changes which the older norm has undergone as it approaches the Calcutta spoken standard. He also notes that, unlike Tamil and other South Asian diglossic languages, literary Bengali (SB or *sādhu bhāṣā*) is rarely spoken. In a context requiring the formal language "SB vocabulary with predominantly colloquial pronunciation and grammar" would be used. (Dimock 1960:44). Compared with other diglossic languages, such as Arabic, which "may be considered in some sense as two different languages, with clear and linguistically definable differences between them" SB and CB (*calit bhaṣa*, colloquial language) "might be better thought of as opposite poles of the same language." (Dimock 1960:44)

> If one considers that the SB and the CB are at the two opposite poles of the same language, it is clear that descriptively there is an infinite number of points between them. Many of these points are occupied by individual writers whose work represents the state of the language at some phase in the history of the movement of the two poles toward each other. . . . If one could graph the accepted norms of the SB and CB over the last century, I suspect the graph would show a truncated pyramid. (Dimock 1960:45)

Thus the two norms, SB and CB, share many features and as time goes on, seem to be approaching more and more a single set of congruent features. At the moment, however, only 39.5% of lexical items are shared, for example. In general, the kinds of changes going on involve such things as simplification of consonant clusters, e.g. SB /smɔrɔn/ becomes CB /šɔ̃rɔn/ 'memory' and morphological changes in verb forms and in certain noun endings, e.g. the expression, "He has given (it) to us" would be in SB /amadigɔke dan kɔriyaɔhe/, while the same expression in CB would have the more simplified form /amader diyeɔhe/. The difference in verb forms (1st person forms for all tenses) and the simplification of the CB forms from SB forms can be seen in Figure 56.

Tense	SB form	CB form
Simple present	coli	coli
Simple past	colilam	collam
Conditional	colitam	coltam
Simple future	colibo	colbo
Pres. continuative	colitechi	colchi
Past continuative	colitechilam	colchilam
Present perfect	coliyachi	colechi
Past perfect	coliyachilam	colechilam
Infinitive	colite	colte
Conjunctive	coliya	cole

Figure 56. CB-SB Lexical Differences (Dimock 1960:47)

Similarly, the changes in nouns from CB to SB form can be seen in forms like the following, where the most obvious difference is that of the case endings when suffixed to the animate plural suffix /-ra/:

Case	SB form	CB form
Nominative	amra	amra
Genitive	amadiger	amader
Objective	amadigɔke	amader, amadike, or amaderke

Figure 57. Difference in Bengali noun endings (Dimock 1960:49)

Other differences include a general abandonment of distinctions of gender, a trend away from the use of compounds, and the elimination of sandhi. Prefixes of quality and negation, such as /su-, ku-/ and /ɔ-/ are now rarely used in CB, since they are more characteristic of SB.

Since the development of CB is linked to the growth of Calcutta and to the role of Fort William College, we see that increased urbanization and contact with foreign cultures, particularly missionaries and their activities, has affected the trend toward the merger of SB and CB. This is obviously not the case in other diglossic South Asian languages, although the influence of missionaries and foreign culture may have had some effect in drawing the attention of some South Asians to their previously ignored indigenous traditions. But the strong correlation between the two in the Bengali situation is apparently unparalleled elsewhere in the subcontinent, and probably makes the Bengali situation a unique one.

The situation in Sinhala-speaking areas of Sri Lanka represents yet another kind of diglossic situation that differs from both the Tamil and Bengali setup. Some of the features of this situation have bearing on our understanding of diglossia in other parts of the subcontinent, even if they are not so characteristic of other areas.

Gair (1968) attempts to provide an overview of diglossia in the Sinhala areas. It begins with a rundown of the various varieties of Sinhala competing in this diglossic situation, then gives an enumeration of the grammatical, phonological, and lexical differences among these varieties, and concludes with a classification of the various varieties of Sinhala into (1) Literary, (2) Formal Spoken, and (3) Colloquial Spoken. The article also notes the existence of regional varieties of Sinhalese, but does not describe them.

In an interesting article, De Silva (1974) explores the influence of linguistic convergence in the evolution of modern Sinhala. He shows that there has been an ongoing tension between the puristic maintenance of high literary varieties of Sinhala and the existence of evolving colloquial forms of the language. He states that the evolution of Sinhala has seen two kinds of puristic traditions, one Sanskritic used in prose, and a non-Sanskritic norm used in poetry. Neither was able to achieve supremacy, and the competition between the two norms gradually lead to the acceptance of fluctuation as part of the scheme of things, and to an on-going hybridization of various varieties. De Silva describes a cyclic situation involving breakdown of one norm, hybridization with other norms, followed by a revival of purism with a return (never quite complete) to older (never quite authentic) norms. De Silva's evidence supports the general contention that diglossia in South Asia does not merely involve tension and alternation between two forms, but a kind of dynamic interaction between them, with both norms influencing each other in a long-term stable relationship.

The evolution of a "triglossic" modern Sinhala from Sanskrit is shown in Figure 58. Unbroken arrows indicate that the varieties referred to are definitely part of the same language; broken lines (- - -) indicate that it is not clear that the varieties are related as in the previous case, and parallel lines // indicate a clear break between the various speech forms. De Silva feels that there are three major characteristics of diglossia as is demonstrated in the Sinhala situation:

Firstly, the maintenance of diglossia (at least in the

> languages under examination) is a puristic endeavour. Secondly, despite purism, the divergent varieties, while maintaining their individual character, tend to converge and generate a multiplicity of hybrid forms under certain circumstances (e.g. popularization of literary pursuits). Thirdly, these hybrid varieties have a tendency to fluctuate between extremes and behave in the way varieties of creoles do in the context of model languages. (De Silva 1974:67)

We would add that while purism is a kind of conscious effort to control the form of a language, there are processes operative in language change that are difficult or impossible to control, and these processes, coupled with populist tendencies, (e.g. in *bhakti* movements) result in a movement away from purist standards. Since those writers whose language is held to be prestigious are not necessarily trained in historical linguistics, especially not during the historical periods De Silva is examining, the linguistic features whose use is advocated by puristic movements are not always etymologically valid historical forms of a language. Certain historical changes may escape the attention of the purists while other changes are cited as examples of the increasing "decay" of the language.

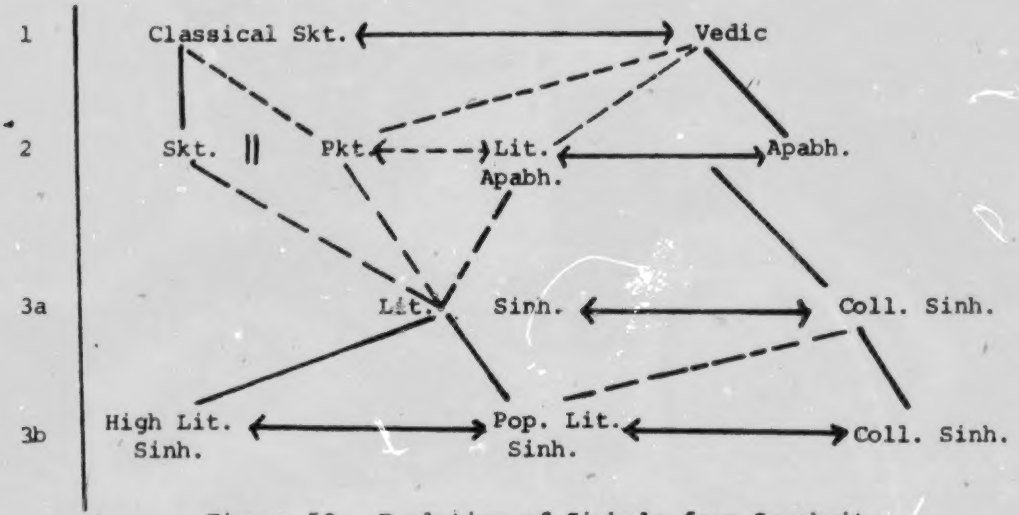


Figure 58. Evolution of Sinhala from Sanskrit
(De Silva 1974:66)

Thus hybrid forms, incorporating features of different stages of the same language, or even of different languages, evolve. By these criteria, even English has to be considered a kind of hybrid, since it has borrowed heavily from other languages (French, Latin, Greek) as well as retaining some features of older English which would have passed out of the language, had there been no attention paid to them--for example, the /ng/ ending of verbs and other words such as 'singing, nothing' etc. became simply /n/ ('nothin', singin') but was reinstated by purists in formal English. Thus practically all English speakers vary between formal /singing/ and informal /singin'/ as two styles in their verbal repertoires.

Another interesting fact about Sinhala diglossia brought out by De Silva is that the formal literary norm, because it can be understood by illiterates, is not perceived by many Sinhala speakers as a different language from colloquial Sinhala. De Silva as part of his study tested the comprehensibility of Literary Sinhala for people with no schooling, and learned that the extra morphological baggage of Literary Sinhala is simply redundant as far as comprehension is concerned. More problematical for mutual intelligibility are lexical differences; if these are not great, the morphological complexity is perceived as just a kind of embellishment of the language and does not impede understanding. This is a finding that needs to be tested in other diglossic situations since it might explain why literary forms of some diglossic languages are accepted without question by the society as being the "same" language as the spoken language, despite great differences. It also offers a clue to proponents of modernization as to how certain compromises between literary and colloquial norms might be made.

Diglossia in Telugu has been examined by Sjoberg (1962). She is particularly concerned with the reasons behind the maintenance of both Sanskrit and Telugu phonological systems in the two norms, and attributes this to the fact that "Hinduism classes as sacred numerous objects that the Westerner would consider elements of the natural world." (Sjoberg 1962:276). This means that these sacred objects have Sanskrit names and consequently Sanskrit phonological realizations are found in everyday Telugu. Thus the Sanskrit norm cannot be relegated to the status of a "restricted literary elite" language. She notes that elites like to use the formal style to isolate themselves from upward-mobile non-elites; the informal style is maintained (for these elites and others) because women traditionally do not acquire learning and by extension, the formal style. Men must thus speak informally to women, children, and other formal style illiterates. She also notes that the informal educated standard is moving forward hand in hand with industrialization and modernization, so that the gap between the informal educated style and the Sanskritized style can probably said to be increasing, in that fewer and fewer speakers feel the need to control the Sanskritized style. Thus we have another example of *distance maintenance* between different social groups being reinforced by the literary-colloquial dichotomy. Those speakers who are unfamiliar with the literary style make no attempt to emulate it, although they respect it as representative of traditional values. Those speakers for whom the literary style is highly valued redouble their efforts to preserve and purify the style to stem the tide of increasingly modernized forms.

5.5. *Studies of cross-language variability and change*

Many studies of social dialects in South Asia have been done by anthropological linguists and others interested in concerns proper to anthropology, such as social change. A number of studies are clearly designed to bolster or support one or more arguments about how social change takes place in South Asia, and whether one group is more resistant or more adaptive to social change than another. Examples of these include Bright (1960a), Bright and Ramanujan (1962), Gumperz (1961), McCormack (1968b), Ramanujan (1968), and Shanmugam Pillai (1968b). In most of these studies there is usually a comparison of some morphological, phonological, or lexical features, of usually the Brahman, Non-Brahman, or Untouchable castes, compared with the literary/historical norm, and some conclusions about innovation and conservatism in various groups. Ramanujan (1968) concludes that Brahman dialects conserve morphological differentiation and contrasts, while non-Brahman dialects generalize paradigms, so that morphologically the NB dialects are innovative and leveling. But phonologically, Brahman dialects in Tamil are more innovative, incorporating phonemic changes from Indo-Aryan and English, which Non-Brahman dialects do not adopt. This analysis is supported by Bright's article on Kannada (1960a), where he concludes that the Brahman dialect "seems to show great innovation on the more conscious levels of linguistic change--those of borrowing and semantic extension,--while the non-Brahmin dialect shows greater innovation in the less conscious types of change. These involve phonemic and morphological replacements." He even ventures to give an explanation of phonetic and phonemic change in these dialects: "The upper class would now appear to originate sound change on the phonetic level; the lower class, imitating this inaccurately, produces change on the phonemic level." (Bright 1960a:425)

We have already discussed the article by Bright and Ramanujan (1962) in an earlier section (5.2). Their article, as we already noted, calls attention to the need for a more sophisticated theory of linguistic change, one which brings into play factors such as the role of literacy, the desire of groups to remain distinct even under centuries of close contact, and perhaps other intangibles yet to be elucidated. As we noted earlier, the South Asian linguistic scene, with its enormous social complexity, probably has something to contribute to a theory of linguistic change based on the primacy of social factors; conversely, the South Asian scene can well stand sociolinguistic variability.

5.6. Other sociolinguistic topics

We conclude this chapter with a discussion of a small number of studies which do not seem to fit under other rubrics. Two of these are studies by Bright (Bright 1966b, 1968) in which he points out that social stratification in South Asian languages may involve also cognitive differences (or semantic structural differences) among the speakers of the competing varieties. He notes, for example, differences in kinship systems (quoting Block 1910) and differences in morphology where high forms of the language may make distinctions (e.g. in the marking of the plural) which are not made in the low varieties (1968). He suggests that it would be "unduly venturesome" to claim that "semantic differences between caste dialects may reflect differences in value system from one caste to another" but that field workers should nonetheless turn their attention to this kind of diversity. (1968:460)

Another study which we include here for lack of a better place for it is Emeneau's monograph on ritual language among the Todas. (Emeneau 1974b) There is little in the literature on the use of ritual language, although some unpublished studies do note it. Emeneau's study is concerned with the hierarchization of language, particularly naming devices, associated with the hierarchization of the Toda dairies.

> In prose there is a large vocabulary relating to the practices of the dairy and all its accompanying operations and objects. But for many of the entities and operations there is a doubling of vocabulary. One set of terms is used only in and of the ti· grade (A), the other set in and of the

> remaining grades (B, C). [See diagram below, Figure 59]. . . and with this other set are to be classed the remaining undifferentiated items; all this set forms part of the ordinary prose vocabulary. . . Apart from this type of ritual utterance, concerned with the details of dairy practice, there is another type which accompanies the practice and which is usually referred to as "prayer". . . It actually consists of two types of utterance. . . . Such prayers, asking for other benefits, are uttered on other than dairy ritual occasions, e.g. at clan or tribal prayer ceremonies. . . ." (Emeneau 1974b:7)

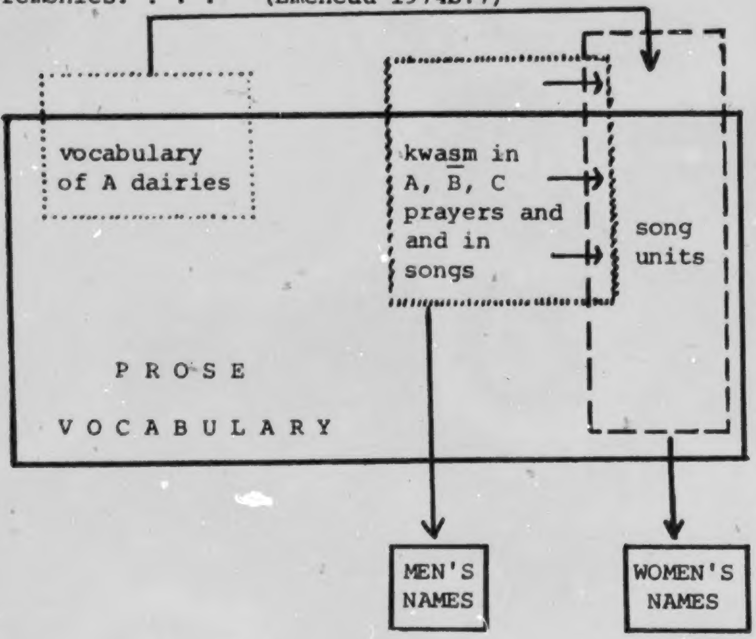


Figure 59. Overlap in *ti·* vocabulary with prose vocabulary (Emeneau 1974b)

What Emeneau seems to be saying is that there is language behavior in Toda which cannot be explained without reference to the culture, especially the ritual. This ritual language is graded in hierarchies, and corresponds to the graded hierarchies of dairies; there is also song language, which overlaps somewhat with prose language; so does the vocabulary of dairies. Names of men in particular are derived from dairy vocabulary in large part. Women's names are not clan-connected in contrast with men's; since men's clans are connected with dairies, they will have names appropriate to their clans.

> The difference in name derivation is a feature of the profound difference in clan status between women and men, as women's inability to use the two kinds of special ritual vocabulary

> in a ritual sense or to have names derived from the kwasm [sacred names] of the ritual language is a feature of their different status from men with regard to the ritual. . . . Frequently it is possible to identify derivation from song-units. . . . Otherwise, . . .there is derivation from the vocabulary of several well-marked semantic spheres, e.g. names of women's ornaments and clothing, characteristic activities of women, hospitality, objects of beauty, and so on. (Emeneau 1974b:9)

Emeneau thus broaches a number of topics not discussed elsewhere in the literature, i.e. sacred or ritual language, naming practices, and in particular, differences in men's and women's naming practices. The lack of discussion elsewhere of ritual language may be explainable by the usual lack of access to it by the profaning presence of the outsider. Emeneau reports that some Todas objected to his being given information of the sort he obtained. But naming practices, in particular naming associated with differences in clan membership and ritual language is a field which could probably be investigated with rich results; the Todas cannot be the only group in South Asia displaying this feature.

5.7. *Conclusions and desiderata*

It is clear from the foregoing that although many studies of sociolinguistic variation have been made in the South Asian context, there is still a fertile field for investigation of the kinds of phenomena now being done in the west. The extant studies reflect concerns current at the time they were written: one notes a reliance on theoretical models such as Bloomfieldian linguistics, which ignored or even scorned any variation at all, or generative models which were interested in finding rules, rather than exceptions to the rules. Anthropological linguistics has always been interested in South Asia and its languages, but often its concerns lie in explaining social structure through language variation rather than language variation through social differences. We would hope that future studies of the South Asian linguistic scene will focus on problems of variation within the speech of the individual, how he or she varies his/her speech according to the social situation; on the educational level of interlocutors, the formality/informality of the situation and the subject matter, and furthermore we hope that these studies will examine the rules underlying this behavior.

We would like to see less of the acceptance of the notion that non-standard dialects have evolved from an older written norm rather than from older unwritten ones. We would like to see more research conducted on bidialectalism and problems of the acquisition of literacy in diglossic situations. There is also a critical need for research on language acquisition and second-language learning problems in South Asia. Substantial progress has been made in some of these areas in the past 15 years, but it is now time for new and bold approaches to old and seemingly pat questions.

NOTES: CHAPTER 5

1. See particularly Bean, 1974, Pandit, 1972b, and Pattanayak, 1975.

2. Cf. Bloch, 1910.

3. "The question arises: is it possible to use models derived from European cases to explain the Indian case? Or, is the caste structure so different from the class stratification current elsewhere as to invalidate the application of models derived from non-Indian society?" (Gumperz 1969b:600)

4. For a discussion of the interaction between objective patterns of linguistic variation and subjective impressions of those patterns, see Wolfram and Fasold, 1974:23-25.

5. Bloch, 1910:1-2.

6. Bloch, 1910:2-3.

7. Bloch, 1910:27-30.

8. Bloch, 1910:5-23.

9. Bloch, 1910:27fn.

10. Bloch, 1910:28.

11. Bloch, 1910:27.

12. Or perhaps, more accurately, of linguistic differences correlated with caste distinctions.

13. McCormack, 1960.

14. "There is also a greater range of variation among NB dialects both

regionally and sect-wise, than among the B dialects. Anyone attempting Tamil dialect-geography necessarily must concentrate on the NB dialects." (Ramanujan, 1968:471)

15. "At first glance, it appears that I [= Iyengar, a Brahman dialect] innovates more than M [= Mudaliyar, a non-Brahman dialect]. But a closer look shows that both I and M innovate (compared to W [=written Tamil]), but in different directions: I toward differentiation, M toward generalization of paradigmatic patterns. In no simple sense is one dialect more "conservative" or "archaic" or resistant to innovation than the other." (Ramanujan, 1968:470)

16. Shankara Bhat nowhere attempts to justify this procedure. Its employment is well demonstrated in quotations such as the following: "Initial *h* of the standard dialect gets elided in both the central and eastern varieties; the west dialect keeps this consonant intact." (Shankara Bhat, 1967-8:68)

17. "The Brahmin dialect differs from the local one in being more similar to the standard language." (Shankara Bhat, 1967-8:71)

18. Cf. Friedrich, 1961.

19. See particularly Harold F. Schiffman and Carol Eastman (eds.), *Dravidian Phonological Systems [DPS]*, "Introduction: Part 3, Verbs." Seattle, Institute for Comparative and Foreign Area Studies, University of Washington, 1975, pp. 240-1.

20. Andronov, 1962:37-8.

21. Shanmugam Pillai, 1968:8-9.

22. See section 2.5.

23. Cf. *DPS,* Introduction, p. xix.

24. Shanmugam Pillai, 1965b:100.

Chapter 6

Individual and Group Linguistic Repertoires

6.0. Bi- and Multilingualism in South Asia

As any visitor to South Asia has experienced, especially one who has attempted to grapple with learning a South Asian language, there is a tremendous variety of linguistic codes in use in the area--not only standard languages with well-developed literatures, but subvarieties that are only spoken and never written; codes used by millions of people over large areas, and codes used only by small communities; codes used only in the home or marketplace, and those used only for writing, schooling, business or religious purposes; codes with long literary histories, and those whose historical origins have never been considered worthy of study; codes with great literary and religious prestige; and codes with little or none. What is more frustrating to the western visitor is that many if not most South Asians control more than one of these codes, and think nothing of switching from one to another in the course of one conversation. The visitor who has learned a variety of, for example, Telugu sufficient to enable him to survive on the streets of Hyderabad will find that that particular code of Telugu cannot (or will not) be used for all the purposes for which the visitor's western language is used at home. The ordinary Hyderabad citizen may use Telugu at home, Sanskrit in the temple, English in the University, Urdu in commercial transactions, and may also control some other varieties of Telugu or perhaps even Tamil, Malayalam or Kannada for reading poetry, for dealing with servants, taxi drivers, or whatever.

In short, the average citizen of a South Asian country as a matter of course has occasion to employ various linguistic codes in the course of his/her daily life, and thinks nothing of this state of affairs. When the history of the subcontinent is taken into consideration, it also seems probable that this kind of multi-code use ("multilingualism") has been characteristic of the region since at least the arrival of the Aryans.[1]

Westerners interested in the Indian linguistic arena have been increasingly interested in bilingualism and multilingualism in recent years, as the topic has come to occupy an important place in Western studies. There are some important differences in the studies of Western bi- and multilingualism and their counterparts with regard to South Asia, and these differences will be brought out in further detail. In fact, it has become clear that the general study of bi- and multilingualism has much to gain from data obtained from South Asia.

6.1. Census Data on South Asian Bi- and Multilingualism

Given the proliferation of autonomous linguistic codes in South Asia, and the widespread confusion over proper label for designating these codes and particular varieties of them, it is not surprising that the primary source for studies on South Asian bi- and multilingualism is census data. The amount of this data is vast and its interpretation poses major theoretical and practical problems. As has been frequently pointed out, South Asian language census returns generally indicate not so much the actual languages or dialects spoken by individuals as the names of the codes *reported* by those individuals as being spoken. In a European context the distinction is frequently meaningless and there is little problem in sorting out speakers of such autonomous codes as English, French and German.[2] In South Asia, however, there are many factors which obscure the validity of the reporting of language data in various censuses. First of all, the names reported by individuals for indicating codes which they use often bear little similarity to commonly accepted linguistic terms. The terms they use may in fact indicate the name of the respondent's region or religious affiliation.[3] In cases where there are numerous terms for referring to the same or overlapping linguistic codes, the choice of a specific term [Hindi, Urdu, Hindustani, etc.] may depend on religious, social, educational, or political considerations.[4] This is complicated by the fact that data computation based on primary census data is, at best, inconsistent and open to political manipulation at worst.[5] Nevertheless, census data as provided in official publications such as Nigam's *Language Handbook on Mother Tongues in Census* (Census of India, 1971) have been the starting place for the serious study of bi- and multilingualism in South Asia.

One of the earliest of these studies (Hodson, 1936) was interested in returns of the 1931 census which indicated that some groups, particularly certain tribes, had significantly higher incidences of bilingualism than did others. Bilingualism ranged from highs of 85% (Saurashtrians in Tamil) to much lower figures, especially for Indo-Aryans in North India. Hodson described how bilingualism may begin very early, as in Nagaland, where children seem to be bilingual from earliest childhood, due to the extreme linguistic diversity in that area. He also pointed out that due to the great divergence between standard Hindi and the dialects of Hindi, many people may be bi-dialectual or even bilingual without this ever being returned in the census, since they call both their local variety of speech and standard Hindi by the same name.

Hodson also attempted to delineate five different kinds of bilingualism, although his taxonomy is based on the notion that bilingualism in two of the major languages of India (e.g. Gujarati and Marathi) is different from bilingualism in two unrelated languages (e.g. Marathi and Kannada) or a major language and a tribal language. Unfortunately, Hodson does not elaborate on what the qualitative differences might be. His article is also concerned about the future of tribal languages, and whether growing bilingualism among rivals may be signaling the end of their own language. He concludes that this is not necessarily the case.

Weinreich (1957) continued the trend set by Hodson in his study of the 1951 census. He was interested in determining (1) why it is that some groups exhibit bilingualism and others not, and (2) what the 'functional load' of a particular language was in a bilingual situation. Hodson had suggested that the social status of the given M(other) T(ongue) group was important in influencing which group became bilingual (i.e. groups would tend to learn a language with higher prestige than their own). Weinreich quotes Opler as suggesting that "it is the same sense of pride and cohesiveness of the community, rather than its isolation, that is at the root of its reported unilingualism." (Weinreich, 1957:213) That is, if a group which we would expect to be bilingual because its own language is not widely known by other groups, nor even spoken by many people, remains monolingual, we must look for other factors such as pride and cohesiveness to explain this. Weinreich points out however, that the incidence of bilingualism is very roughly inversely proportional to the *relative size* of the MT group. He devotes the last part of his article to computing indices of 'functional load' of the languages MT and O(ther) T(ongue), which is to show that a language which may not have many speakers may have a high functional load through bilingualism--e.g. English exhibits higher functional load, perhaps even higher than Hindi, although Hindi is spoken by more monolinguals than is English. Also Kannada is functionally very important in South Kanara district (Mysore State, now Karnataka) because almost everyone there has a MT other than Kannada but only Kannada has official status. Yet the population figures for South Kanara district are not high. Unfortunately, computations such as Weinreich's must be taken with a grain of salt due to the inherent unreliability of census data.

Weinreich's researches into Indian bilingualism are continued in Davidson (1969). He states that it is difficult to determine the importance of English as a bilingual's OT because of apparent attempts to suppress certain data by its presentation in the census; nevertheless, in Davidson's survey English emerges as a very important OT (in Weinreich's sense, with high functional load), surprisingly even in Uttar Pradesh where anti-English sentiment is high. Hindi, of course, remains an important OT in certain other states, mostly northern.

The only other country in South Asia where census bilingualism studies have been done is Sri Lanka (Coates, 1961). This study is mainly statistical and draws no great conclusions except for noting the decline in the reported use of English between 1946 and 1953.

6.2. General and Theoretical Studies of Bilingualism

As mentioned earlier, there are many studies of bilingualism in the west, both older and modern, that are of some interest to us because of what they can explain about the importance of multilingualism in South Asia. The literature is abundant with case studies of individual situations like French-English bilingualism in Montreal, French-Dutch bilingualism in Belgium, Norwegian-English bilingualism in America (Haugen, 1953) as well as many articles of a theorectical nature which have not taken the subcontinent into account and probably should have. Individual case studies from the subcontinent will be dealt with below.

We will consider here the general studies done usually in the west which have some bearing on multilingualism in South Asia, ignoring those studies which lack relevance to the subcontinent, or are repetitive applications of one or another theory of bilingualism to a given area, even though some of them might replicate conditions found in one or another South Asian subgroup.

The first problem in dealing with the topic of bilingualism in general is definitional. Haugen (1953:7) defined the lower threshold of bilingualism as beginning "at the point where the speaker can produce meaningful utterances in the other language." Haugen also later (1956:10) observed that there might be a number of dimensions to bilingualism, with gradations of mastery, and noted that the amount of linguistic "distance" between languages is an important factor which can influence the amount of bilingualism in the language. Diebold (1961) challenged Haugen's *minimal* threshold for bilingualism (complete meaningful utterances) since his study of Huave in Central America showed that although they were unable to produce meaningful utterances, they nevertheless showed the influence of Spanish in their Huave, and if one ignored these early stages, one could not study the influences of one language on another, as the people begin to become bilinguals. What is interesting to note in Diebold's study is that he notes *variation* in the shape of borrowed items, e.g. Spanish '*hasta*' 'until' may be /asta, ista, ísta/, or /ast/ in Huave, with no telling as yet which form will win out. Diebold proposes a measure of incipient bilingualism based on the use of the Swadesh lexico-statistics dating list, and offers a modified definition of minimal bilingual skill: "contact with possible models in a second language and the ability to use these in the environment of the native language." (Diebold, 1961:111) For South Asia, it is clear that what has been studied is mainly more complete bilingualism, or interference in one language from another where bilingualism is common.

Interference is another phenomenon noted in bilingualism, defined as "deviations from the norm of either language which occur in the speech of bilinguals as a result of their familiarity with more than one language" (Weinreich, 1953). As Haugen notes, (Haugen, 1956:12) "The identification of interference can be done with certainty only when we have a base line from which to start, a state of language immediately preceding the bilingual event." As we have noted, the "base line" of lack of contact may be indeterminable in South Asia for all practical purposes.

Although the distinction between the bilingual individual and the bilingual community was made by Haugen (1956), Fishman (1967) attempts to construct a theory about societal bilingualism and its stable and transitional aspects, dwelling on the diglossic features of some bilingual contexts. The emphasis which Fishman places on the qualitative difference between individual and societal bilingualism is quite necessary, and the distinction between the two types of bilingualism is quite useful in the South Asia context.

Another contribution to the literature on bilingualism which has some bearing on South Asia is the proceedings of an international seminar held in New Brunswick, Canada (University of Moncton) in 1967 (Kelly, 1969). It contains a number of useful studies and discussions on a wide variety of topics concerned with attempts to define, measure, evaluate, and delineate bilingualism and its acquisition, proficiency, effects, roles, behavior, incidence and distribution. One of the important points of this book is the necessity of distinguishing *personal* and *institutional* or *official* bilingualism, i.e. a state of affairs where members of a society may not be bilingual to any extent, but the state or its offices and organs may be. In Belgium, for example, very few are bilingual, but the state is completely so; in South Africa, in contrast, individual bilingualism seems to be very high (above 60%) but the amount of official bilingualism is not as noticeable as in Belgium. In South Asia, of course, figures are harder to assess because of the difficulty of interpreting census materials, but it is clear that any future study of bilingualism in the subcontinent would benefit greatly from an application of many of the principles and suggestions laid forth in this work. One example is the attention paid to kinds of measurement techniques, e.g. in the article by MacNamara (MacNamara, 1969) where the efficacy of various techniques of measuring bilingual proficiency are discussed. He also points out the importance of delineating the *context* of bilingualism, i.e. whether a person learns a language in school, at home, or on the street, and how it is reinforced. One must be sure of what one is testing, since different kinds of tests test different things. It is clear in comparing MacNamara's discussion of kinds of tests with the work done on South Asian bilingualism, that few of the latter are reliable in their results, because of the lack of sophistication of testing methods employed.

Another important article (Mackey, 1965) attempts to illustrate how specimens of bilingual behavior may be analyzed and measured. He notes the great incidence of *variation* in the speech of the bilingual, far beyond what is normally permitted by either of his or her languages.

The possibility of classifying multilingualism in its various modes with a greater degree of sophistication is provided in Stewart (1962b). He sets up a code by which one can refer to different languages used in multilingual situations: first the type of language (capital letter) then its function (lower case). Thus, S(tandard), V(ernacular), K(Creole), C(lassical), A(rtificial), P(idgin) and M(arginal) languages can be combined with o(fficial), l(iterary), g(roup), or r(eligious) functions. He also mentions some language attitudes which are involved in the typology, e.g. historicity (whether a language has developed through use), standardization (whether one or more norms are codified), vitality (whether language has a community of live speakers), and homogenicity (whether the basic lexicon and grammar are derived from the same prestages of the language--Creole and Pidgin languages obviously do not fit the criterion of homogenicity). The terminology provided by Stewart has obvious applications in South Asia. Thus, for example, one can use it to differentiate between a multilingual person who uses Hindi as a Sg (Standard language, group function) from one who uses it as a Kg (Creole, group function).

6.3. Studies of Bilingualism in South Asia

6.3.1. Historical

The earliest studies of bi- and multilingualism in South Asia were concerned with the *effects* of bilingualism, e.g. in the creation of linguistic areas (Emeneau, 1954), or in the history of Sanskrit after its speakers entered India. Emeneau assumes, as do most other writers, that extensive bilingualism was the cause of the changes in Sanskrit, e.g. the emergence of retroflex consonants where there were none earlier was due to the existence of many Dravidian speakers who were bilingual or imperfectly bilingual in Indo-Aryan languages and their own. In another article, Emeneau (1962a) focuses on Brahui (Dravidian) and Balochi (Iranian) and shows that not just bilingualism, but extensive *bilateral* bilingualism is responsible for the great amount of structural borrowing between the two languages. Emeneau specifically condemns the search for substratum explanations for structural borrowing without relying on adequate data, since the substratum explanation is speculative and becomes a catchall or a "wastebasket" for anything unexplainable by hard data. Emeneau summarizes work previous to his own as exemplified by many of the papers in Ferguson and Gumperz (1960), and asserts that these papers make two basic assumptions: (1) that the solidarity of a separate social group with a language community will find expression in differentiating linguistic features (groups will use language differences to express their social differences) and (2) that a speaker with prestige is imitated linguistically by other speakers with less prestige. The conclusion is that as old features disappear through imitation, new ones will appear through innovation (or else there would remain no linguistic diversity after everyone imitated everyone else). But as Emeneau shows, this does not explain how it is that some groups do *not* imitate other groups more prestigeful than themselves:

> It must be emphasized that in many respects the system exerts pressure to preserve differences, rather than to eliminate them. . . . The existence of three mutually unintelligible Dravidian languages in the Nilgiris after many centuries of coexistence by three comparatively small communities in a small isolated area, can only be explained in terms of preservation of difference. . .with a castelike, heirarchial structure of three communities. (Emeneau, 1962a:432)

In other words, South Asian communities value their differences and these differences are reinforced by the caste system. Thus the prestige factor, whatever its explanatory power may be elsewhere, has been exagerated. Emeneau states that "it is necessary to emphasize that the Hindu and Hindu-like social structures of the Indian subcontinent may militate in many cases against the pressure of prestige operating in the direction in which it is assumed that it does in the West." (Emeneau, 1962a:432) By way of support Emeneau cites the example of Latin spreading at the expense of many Western European languages, but Norman French never getting a foothold in England. In relating this example to South Asia, he notes that "it is hopeless to think of knowing, without any direct evidence at all, why Sanskrit came to supersede the aboriginal languages of North India." (Emeneau, 1962a:433). Emeneau considers it to be a commonly admitted doctrine that "extensive borrowings from one language into another can only occur through the agency of a bilingual section in the joint community." (Emeneau, 1962a:434) Because of this he asserts that the fact of Sanskrit borrowing heavily from Dravidian (and other aboriginal languages) both in vocabulary and structure can have happened only through bilingualism. He notes, however, that "nothing is known of the Indian social and political structure into which the Sanskrit speaking invaders made their way. . ." (Emeneau, 1962a:434).

Emeneau then concentrates his attention on Brahui and Balochi, where an almost unique situation of *bilateral bilingualism* exists, which he claims has been ignored in the earlier literature on bilingualism. He notes that "it may even be that in the general examinations of the problem there is a hidden, ignored assumption that bilingual community situations are always unilateral; generalization without recognizing this assumption may well have invalidated or weakened some of the general discussions of this topic." After reviewing the study of linguistic diffusion between families since 1906, Emeneau concludes that "at one time in the history of the Brahui Confederacy there must have been more non-native speakers of Brahui" whose mother tongue was Balochi than there were speakers who learned Brahui from native speakers of the Brahui. Thus the legacy of people learning an OT from people whose MT is not the OT was passed down and represents an important aspect of multiple use of linguistic codes in South Asia--the example of Indian English is the most obvious to westerners. Emeneau concludes:

> The Indian evidence for structural borrowing should now join that from Europe to silence the mid-nineteenth century dogma and Sapir's qualms. 'Language mixture,' i.e. structural borrowing is not a monstrosity or an impossibility. It occurs. There is surely much more evidence of it to be recognized and added to the small amount of certain instances that we may now operate with. To be sure, the only really valid evidence is that derived from bilingual situations in which the languages on both sides are well known. It will not do to deal in substrata that have long vanished entirely from our control. (Emeneau, 1962a:441)

Thus Emeneau comes down strongly on the side of bilingualism, especially *bilateral* bilingualism, as the main vehicle for structural borrowing in South Asia, or elsewhere for that matter. And of course this means that comparisons of two systems means knowing their structures and their prehistories thoroughly, in order to avoid speculation and conjecture.

Another major work in which attention is paid to the historical aspects of multilingualism in India is the compendium of articles edited by Southworth and Apte (CCSAL). In the introduction to that volume, and in an article by Kuiper (1967) bilingualism from an early period is credited for the extensive convergence of phonological, grammatical, lexical and semantic systems of South Asian languages: "Kuiper and others have maintained, for example, that the retroflex-dental contrasts in Sanskrit goes back to the Vedic period, and has resulted from contact with Dravidian languages. Since the frequency of retroflex consonants in Indo-Aryan languages increased with time, the present state of affairs would seem to be a result both of the initial contact and the continuing contact between the Indo-Aryan and Dravidian speakers." (Southworth and Apte, 1974b:13) A previous article by Southworth (Southworth, 1971) discusses two different kinds of bilingualism which might have been involved in the evolution of Marathi: (1) either a situation of what Southworth seems to prefer to avoid calling bilingualism, wherein lower caste Dravidian speakers pidginized the Maharashtrian Prakrits, and thereby developed a new pidgin which eventually was also adopted by higher caste speakers and developed into the forerunner of modern Marathi, or (2) "a different kind of bilingualism, involving fuller control of two languages on the part of a substantial segment of the population." (Southworth, 1971:270) "The validity of this alternative would depend on evidence showing that this type of bilingualism has produced, or does in general produce, the kinds of results presented. . .above." Southworth seems to feel that situation (1), which he avoids calling bilingualism, but characterizes as ranging "from near-native control (in cases of intimate contact) to a true pidgin (which would be the most likely result in the case of the lower-class individuals)" is responsible for the evolution of Marathi. Since there is evidence from the present situation of such a large range of abilities, it seems obvious to Southworth that such a variety probably existed earlier, during the evolution of the forerunners of Marathi. Southworth thus favors a kind of multilingualism involving a spectrum of competences, ranging from native-like control of more than one language to reduced repertoires, like perhaps the incipient bilingualism described by Diebold. He thus differs from Emeneau (1962a) who favors bilateral bilingualism as the important avenue for structural borrowing. Perhaps the latter is necessary for the kinds of parallelism found between Balochi and Brahui; probably the former is more characteristic of Marathi which displays a whole range of social varieties of speech, from very Dravidian-like varieties in lower caste dialects, to more Indo-Aryanized varieties among the higher castes. Southworth's article is probably one of the most important in the area of sociolinguistics to have appeared in the last decade, at least; we will return to it later in this chapter.

An important recently published article concerning the origins of bilingualism in South Asia is that of Nadkarni.[6] This article describes how the formation of relative clauses in that variety of Konkani spoken in Kannada-speaking areas (Karnataka Saraswat Konkani, or KSKo, speakers of which are all bilingual in Kannada) has been influenced by the Kannada structures not found in Indo-Aryan, this leading to the increasing Dravidianization of KSKo. KSKo historically possesses a fully developed Indo-Aryan system for forming relative clauses of the sort demonstrated below:

KSKo jo mhāntāro pepar vāccat āssa(-ki) to ḍākṭaru āssa.
which old man paper reading is that doctor is.
"The old man reading the/a paper is a doctor."

which is similar to the structure in Hindi and other Indo-Aryan languages:

jo būṛhā akhbār paṛh rahā hai vo ḍākṭar hai.
which old man paper reading is that doctor is.
"The old man reading the paper is a doctor."

In spite of having this construction, however, the most frequent kind of relative structure seen in KSKo is of a more Dravidian sort, and does not resemble the relative clause structure of other Indo-Aryan langauges:

Kannada pēpar ōdutta idda mudukanu ḍākṭaranu iddāne.
KSKo pepar vāccat āssillo mhāntāro ḍākṭaru āssa.
paper reading being old man doctor is.
"The old man who is reading a paper is a doctor."

Unlike the Indo-Aryan relative clauses, which can be extraposed, Dravidian (and KSKo) relative clauses of the type just mentioned cannot be extraposed. Thus the KSKo/Kannada type not only share a basic syntactic construction, but a set of restrictions on the operations which can apply to the construction. To Nadkarni, the borrowing of the Dravidian relativization system by Konkani seems to be totally unmotivated, as Konkani had a perfectly adequate system that it inherited from Indo-Aryan.

Nadkarni offers a number of explanations why this kind of borrowing has probably gone on, and they are reasons which we feel are worth noting. Nadkarni points out that restructuring of the syntax at such a profound level is usually a result of pidginization and/or creolization; yet, as other scholars have also noted,[7] there is no evidence for such processes in this area. Another feature of this situation is that bilingualism is not bilateral here, but one-way: only Konkani speakers learn Kannada. Kannada speakers in this area never learn Konkani. Thus we cannot look to a substratum of Dravidian speakers who have brought the Kannada structure into Konkani through their bilingualism, as in other situations we will describe below. Another usual explanation for this kind of borrowing is through code-switching. But Nadkarni has never observed code-switching between Kannada and Konkani in the KSKo community. Code-switching is observed between English and Konkani, since English enjoys more prestige than Konkani. But Kannada enjoys no prestige among KSKo speakers; they learn it only because it is the dominant local language.

In fact the prestige factor is rather complex in this situation, and bears some looking into. For Konkani speakers, Kannada lacks prestige because the KSKo community are Brahmins and feel themselves to be "superior" to the local people. Yet for local Kannada-speaking people, Konkani is of little value or prestige. Therefore while Kannada lacks prestige for KSKo speakers, it has *functional dominance*[8] over Konkani in this area, and the bilingualism of Saraswat Brahmins is a consequence of it.

> To sum up, we have here a clear instance of structural borrowing from the language of a socially less prestigious group into the language of a socially more prestigious community, occurring in the absence of such factors as code-switching, substratum influence, or pidginization and creolization. Nor can the borrowing be explained in terms of the usual linguistic reasons: it fills no structural gap in the language, nor does it equip the language with a wider range of stylistic choices. How then do we account for this phenomenon? The explanation probably lies in the *INTENSIVE* and *EXTENSIVE* bilingualism of Saraswat Brahmins in a region where Kannada is the functionally dominant language.[9]

Nadkarni defines *extensive* bilingualism as bilingualism that "is co-extensive with the entire community"[10], while *intensive* bilingualism is a kind of bilingualism where a speaker of one language is not merely conversant with another language, but uses it for many purposes in daily living. Nadkarni feels that extensive bilingualism is necessary for structural borrowing to become stabilized, since the whole community is susceptible to borrowing, and no one speaker notices anything strange about a feature of another language. Intensive bilingualism, on the other hand, lays the groundwork for the borrowing, since all speakers are constantly using the other language, and want to lessen the psychological load of having two different systems in their heads. They lessen it by allowing the systems to merge, and structures from one to be used in the other system.

Nadkarni also raises the issue of why Kannada structures are filtering into KSKo but not vice-versa. He answers this by explaining that KSKo speakers learn their Kannada in each generation, not from KSKo/Kannada bilinguals, but from Kannada speakers, and eschew the local varieties of Kannada for a more literary standard. Thus any non-Kannada features of KSKo Kannada are not passed on to future generations, but remain idiolectal with a given speaker. They avoid the local varieties because of a desire to not be identified in their Kannada with any lower caste community in the area--a kind of linguistic caste avoidance which seems to insure that the distance between KSKo and Kannada will be strictly maintained, or at least that the Kannada will remain distant from the Konkani, and that only the Konkani will approach the Kannada. Since KSKo speakers favor education, and this has been until recently almost exclusively through Kannada (Konkani not being used as a literary language in this area), children easily acquire a bookish variety of Kannada at school.

But the Konkani spoken in the area, because of this very lack of standardization, is more susceptible to influences through bilingualism, and through transference from previous generations already affected by Dravidian structures. Thus the increased Dravidianization of KSKo is apparently assured. Nadkarni thus is proposing another explanation for the kind of structural borrowing based on the case of the Karnataka Saraswat Konkani speakers of South Kanara. He rules out some of the traditional explanations, such as the prestige factor, and shows that Konkani is maintained through a desire for distinctness noting that imperceptible changes in KSKo are also creeping in through intensive and extensive bilingualism.

6.3.2. Case Studies of Individual Communities

6.3.2.1. Bilingualism as a Burden

Aside from carrying on the study of bilingualism in South Asia as a historical phenomenon which has lead to extreme convergence and areal congruence, recent scholarship has produced what we might consider case studies of bi- and multilingualism in the subcontinent. There appear to be two kinds of case studies in the literature, parallel to the two different approaches found in studies in the West. One approach, typical of the older western literature assumes that bilingualism is a kind of burden on the social body, and introduces problems which make for slow learning at school, decreased social mobility for those characterized by it, or leads to discrimination against these individuals.[11] The second approach to bi- and multilingualism in South Asia assumes the phenomena to constitute an elaborate, but nevertheless integrated, repertoire of linguistic codes in which all of the components fit well together and offer a varied and often fascinating display of linguistic behavior.

Among the "social burden" kinds of studies done on South Asia, a few concentrate on the problems occasioned by the emphasis on English or another medium of instruction which is not the mother tongue of a student, and how this may or may not retard school progress. Chickermane (1971) examines bilingual areas on the Maharashtra-Mysore border, where students may speak Konkani, Kannada, or Marathi, but are in schools where their mother tongue is not used. He begins by classifying bilingualism into three types: type A, where both languages can be used in home, school and in the child's social environment; B1 type, where the child uses one language at home, but a different language at school and in his environment outside of the home; and B2 type, where home language is the same as the environment language, but the school language is different. He concludes that (1) a pure bilingualism of the 'A' type leads to "no significant difference in the achievements of children grouped on a language basis," and that bilingualism is therefore not a handicap for those children. Chickermane also notes that 'B1' bilingualism seems to have no adverse academic consequences. 'B2' bilingualism, however, does, and Chickermane notes a difference at the primary level between Marathi-Marathi groups and Kannada-Marathi groups (the latter are in effect the B2 groups, the former B1) although the difference levels out at the higher primary level. Chickermane speculates that in B2 bilingualism the school language is not reinforced by the environment language--it is encountered only at school. This is seen in the case of Konkani speaking children, in Konkani areas, being taught only in Kannada, but who never have their Kannada reinforced by the environment. This is also the case of Goan children who are taught in English from 4th standard on. Chickermane recommends remedial measures in B2 areas, such as having bilingual teachers who can teach the new language through the medium of the old, and curriculum enrichment.

A very polemical, anti-colonial attitude toward bilingualism involving English is demonstrated in Pieris, 1951. He coins the term "cultural marginality" to describe a person who, through having been educated in English, knows his own language poorly but who also fails to attain full proficiency in English. Pieris seems to consider the resultant condition to be one of hopeless worthlessness and sloth. It would be hard indeed to prove that bilingualism leads automatically to such an imagined condition, but in the early post-colonial period it seems to have been fashionable to see bilingualism as a kind of vestige of colonialism. The only real value of the Pieris article is in providing some interesting examples of Ceylon English. Many more examples of this kind of attack on English as a medium of instruction can be found in the non-scholarly literature, especially in India.[12] Such attitudes unfortunately provide a distorted and biased backdrop to the study of bilingualism and its social effects.

A third example of the "social burden" type of study of bilingualism is Ross, 1965. She studies students in Bangalore colleges who are multilingual in various combinations of languages drawn from Kannada, Tamil, English, Konkani, Urdu, Malayalam, and Telugu. Her concern is with the social utility of language; she seems to think that people learn a language because it is useful (i.e. there are incentives) for learning it, and that usefulness may be contradicted by national sentiment--i.e. it may be more useful in some African countries to learn English or French, but national sentiment induces people to learn a "useless" language other than the colonial ones. Ross' premise is that prestige is the primary factor in causing individuals to require a second or additional language.[13] With this understood she then addresses the "social problems of a bilingual person" in three areas: family life, education, and social life. Her study shows that students from rural areas, where English is not well known, have great trouble in the first year or two of college, where they are forced to learn English or fail; most fail. After that they are more likely to be successful.

6.3.2.2. Bi- and Multilingualism as a Unified Elaborate Repertoire

One of the most prominent advocates of the second approach to the study of bi- and multilingualism, namely that which assumes that bi- and multilingual competence is a type of "expanded" linguistic competence having a structure of its own, is John Gumperz. One example of his approach is an article on Hindi-Punjabi code-switching in Delhi (Gumperz, 1971). In this article, Gumperz reiterates the distinction between individual or isolated bi- and multilingualism, and community bilingualism, where the norms of the separate codes are different. For the isolated bilingual, the norms of the two codes are those of the separate communities; for the bilingual in a situation of community bilingualism, speakers "tend to create their own norms which are quite often different from those prevailing in the respective monolingual societies."[14] For example, English in India "will thus deviate considerably from the norms current among native speakers of English" in other places. "This kind of deviation represents not a failure to control English, but a natural consequence of the social conditions in the immediate environment in which Indian English is spoken."[15]

But there are conflicting tendencies in multilingual societies: "The need for frequent code-switching on the part of a large number of individuals tends to reduce the language distance between codes. . . Linguistic overlap is greatest in those situations which favor intergroup contact."[16] But there is a need to maintain some symbols of role specificity and this is reflected in *deterrents on excessive borrowing,* which keeps the codes somewhat separate, i.e., prevents complete merger. Gumperz then notes the possibility of there being many different styles of the same language, each situationally determined--styles differing in pronunciation, lexicon and grammar.

In this important article Gumperz also examines the phenomenon of Hindi-Punjabi code-switching in Delhi, where code-switching is completely normal for Punjabi speakers. Informants were asked by Gumperz to imagine different contexts and tell how they would respond linguistically.[17] Gumperz found that the Punjabi of Hindi-Punjabi "bilinguals" took on many "Hindi" features while still retaining a number of distinctive features of P, e.g. P /kii/ instead of H /kyaa/ 'what?' and P /nd/ instead of H /t/ imperfective participal markers. Older speakers of Punjabi from the Punjab considered this kind of P to be "bad Hindi," but the Delhi speakers had no consciousness of its being distinct from standard Punjabi. This preservation of a few minimal P items evidently "suffices to preserve the necessary minimum of symbols of role specificity."[18]

In another article Gumperz and Wilson (1971) examine a multingual situation on the Maharashtra-Mysore border (specifically in the Sangli disrict, Maharashtra), where the local dialects of Kannada, Marathi and Urdu have converged into a sort of phonetically and syntactically identical code, where only lexical items and morphophonemic rules differ.[19] Since we have no evidence of pidginization having occurred, but only convergence over the centuries of the three systems, we seem to have here a kind of creolization without pidginization. Thus the authors are able to say that "there seems to be no reason therefore to draw an *a priori* distinction among pidginization, creolization and other diffusion processes; the difference may be merely one of degree."[20]

Gumperz and Wilson show that movement or convergence in this area is toward Kannada and/or Marathi, with Urdu never the model unless Marathi also shares the phonological, morphological, or syntactic feature. Thus highly marked features of various sorts are levelled (although some highly marked categories such as the inclusive-exclusive distinction in the first person plural pronouns are kept, which standard Kannada doesn't have, although older Kannada did).

This paper, like Southworth, 1971, is important because of its recognition of the relationship between bilingualism and creolization. It also discusses why this centuries-old bilingualism/multilingualism should persist; here the difference in religion may be crucial: Kannada speakers are Jains, Marathi speakers are Hindus, and the Urdu speakers are Muslims. By maintaining a single surface (and underlying?) code, but different lexical realizations, they can pretend to maintain their familiar differences while still communicating with each other in a system which shares all subsystems except lexical.

In a recent paper P. B. Pandit (1972) examines "Tamil-Saurashtri gramatical convergence" with a view to establishing common sociolinguistic traits in the Indian area which have arisen due to bilingualism. For a case study he has chosen the Saurashtris, speakers of an Indo-Aryan speech form, originally a dialect of (Old) Gujarati, which has converged grammatically and phonologically with Tamil due to four centuries with the latter in Tamilnadu. All Saurashtris except for very small children are bilingual in Tamil; their language receives no reinforcement from outside the area because it is spoken only in the South (having diverged from Gujarati many centuries ago). "Tamil-Saurashtri bilingualism is. . .a case of linguistic convergence under conditions of stable bilingualism."[21] Their language is also rarely written, since they have no alphabet and the Tamil writing system is not adequate for writing Saurashtri; most Saurashtris are literate in Tamil if anything. Thus the influence of Tamil has been heavy, and the language seems to be closer to Tamil than to Gujarati, except for lexemic resemblances to the latter. Pandit seems to feel that the "true bilingual; the person with a completely separate set of codes at all levels, may exist only in imagination,"[22] and that Tamil-Saurashtri grammatical convergence is a kind of diglossia, with one language used in one set of circumstances and the other used in other set(s). In fact, it seems to us that given the diglossia already existing for Tamil, the situation would be at least triglossic, with Saurashtri used in the home, spoken Tamil in the street, and literary Tamil in writing.

An important aspect is his emphasis on the structurally determined aspects of bilingual linguistic behavior:

> Whether it is stylistic variation among the varieties of one language or whether it is code-switching across mutually unintelligible varieties, *variation is rule governed by behavior* and the analyst has to bring out the complex interplay of this patterned behavior. The models of description of monolingual communication, contrastive interference or translatability, are not suitable because they are based on the assumption that the two languages are distinct at all levels, while in fact convergence of the different varieties in bilingual communication has been frequently noticed; Gumperz rightly observed that language distance is not an absolute; it is a function of intensity of contact and social context.[23]

By this Pandit presumably means that language distance between two genetically unrelated languages is not fixed, and that although for standard Marathi and standard Kannada may be genetically distant, they may have varieties which approach each other under intense contact; so may Tamil and Saurashtri; in this case, Saurashtri has obviously approached Tamil, which has not approached Saurashtri to any measurable extent.

Another example of the convergence-through-bilingualism or through linguistic contact is a study of Emigrant Sindhi and Kacchi by S. K. Rohra (1971). In this short study, E(migrant) S(indhi) is compared with K(acchi) and S(indhi); both ES and K have undergone changes; the former in only 17 years has replaced fricatives with stops and has restricted the occurrences of some short final vowels: K underwent the same rules at an earlier time. Rohra proposes that K is probably an earlier version of emigrant Sindhi, its speakers having emigrated after the 14th century A.D. from Sindh and that it arose when speakers of Middle Sindhi came into contact with speakers of Middle Gujarati, just as modern ES speakers have come into contact with speakers of Hindi and Marathi, which lack the features lost from Middle Sindhi and Modern Sindhi. Since modern speakers are now largely bilingual in some other Indo-Aryan language, he attributes the changes in ES to the fact of bilingualism, and proposes that the same state of affairs (bilingualism of Middle Sindhi speakers with Gujarati) is responsible for the emergence of Kacchi.

Similarly, Upadhyaya (1971) demonstrates that the Bidar dialect of Kannada has been highly influenced by Urdu because of the area being under the jurisdiction of the Nizam of Hyderabad for a long period. He enumerates many lexical, phonological and syntactic changes, the most interesting of which is perhaps the quotative sentence types with the subordinate clause following the "X said" clause, and an Urdu quotative particle /ki/ added: /avə̃ anda ki na:bhi: nim sari barte/ instead of standard Kannada /avanu nim joteeli bartiini anta heeLidanu/ where the subordinate clause precedes the sentence-final verb, with a Dravidian quotative marker /anta/ instead of /ki/. Other interesting features of Bidar Kannada are the way borrowed verbs are incorporated into the lexicon by using the /isu/ suffix added to Urdu roots, by extensive use of /maaDu/ 'make, do' /aagu/ 'become' or /haccu/ 'attach'; the use of the Urdu genitive suffixes /kaa, kee, kii/, the borrowing of numerals from Urdu and Marathi; the phonological contrast between /a/ and schwa; phonemic nasalization and borrowing of aspiration; dropping of final vowels to result in word-final consonants (whereas standard Kannada and Dravidian languages in general prefer utterance-final vowels), and others.

It is obvious from the foregoing studies and others that bilingualism is responsible for a great deal of convergence in structure found in the languages of the subcontinent. We have dealt here with convergence where the final product is still felt to be a variety of one language, rather than a completely new language or creole; yet, as some have pointed out, the distinction between creolization and convergence is not necessarily a clear one. In the next section we deal with studies of creolization and pidginization per se in the subcontinent, and will also deal with South Asian English and the result of its contact with the language of the area. In conclusion we will discuss how convergence through bilingualism and multilingualism differs, if at all, from creolization and pidginization, and how studies of both topics in other parts of the world might benefit from a closer look at the

Indian subcontinent.

6.4. Pidginization and Creolization

6.4.1. General Theories of Pidginization and Creolization

As interest has developed in recent years in Creole and Pidgin languages, focussing primarily on their African and Caribbean varieties, but with some attention to Asia too, there has been a refinement of the theoretical concepts developed to handle these varieties of language. It is not possible to review all these theoretical developments in extenso here, but a number of directions will be mentioned and the main theoretical thrusts which have any bearing on the topic of creolization and pidginization in South Asia will be outlined. A very concise and useful introduction to some of the latest thinking that is in print can be found in Hymes, 1971.[24]

A number of definitional problems exist with regard to pidginization and creolization, and scholars working in this field do not as yet fully agree on some basic definitions of these terms. It is at least accepted, however, that *pidginization* refers to the process or set of processes leading to the development of a *pidgin*, this being a reduced version of some language used for trade or other communication in a situation where the *mother tongue* of the speakers involved is some other natural language. That is, a *pidgin* is not commonly held to be the native language of its speakers. It may arise anywhere where people of various social groups in contact have no language in common. It arises out of the immediate need to bridge this communication gap; it may survive only a short period or endure for centuries. When two speakers of a pidgin have offspring who grow up with no other language than the pidgin as a mother tongue, their mother tongue is called a *creole*, and the process of the development of the pidgin into a creole is called *creolization*. Creolization is generally taken to refer to the expansion of the limited, reduced pidgin into a full-fledged language, capable of expressing whatever its speakers wish to express. This commonly takes place in the first generation that the creole exists *qua* creole, although of course further developments may alter the development of the creole in successive generations, it being a natural language and subject to all the processes that natural languages undergo. Were this process not to be completed in the first generation, the speakers would supposedly be speaking a reduced form of a natural language, a theoretical prospect which most linguists would reject.[25]

One of the processes involved in the evolution of creoles and pidgins is *relexification*, whereby the vocabulary or lexicon of one creole/pidgin is replaced somehow by a different vocabulary from another language. For example, it is Hesseling's contention (Hesseling, 1899) that Afrikaans began as a Portuguese pidgin (with a great deal of Malay vocabulary) used by the 'coloreds' of Capetown. Then this pidgin was relexified with Dutch vocabulary and was creolized, becoming the mother tongue of Afrikaaners and eventually acquiring status as a full-fledged language. This process of relexification may be fairly abrupt, or may take generations. Whinnom (1965) claims that most of the pidgins of the Orient, because they demonstrate many shared structural properties that could not have arisen due to chance, were derived from Portuguese-based pidgins that have been variously relexified, drawing vocabulary from different prestige languages (i.e. English, Spanish, Chinese). These creoles are believed by Whinnom to owe their origin to a Portuguese pidgin brought by the first Portuguese mariners.

Another topic which occupies the attention of creolists is the question of genetic origins and classification of languages. A fundamental question of creole studies is whether all pidgins/creoles can ultimately be derived from a proto pidgin, thought by Whinnom and others to be a Portuguese pidgin descended from medieval mediterranean Sabir, or whether the great similarities found among creoles throughout the world can be attributed to some universals of language that naturally find expression when systems are brought into contact. Both of these positions, called neogenesis and polygenesis respectively, have numerous advocates. This question of course interests South Asianists, because while some obviously Portuguese-based creoles are found in the subcontinent (Schuchardt, 1889; Dalgado, 1900; Fonseca, 1959; Thompson, 1959; Theban, 1973;), others with no European content, such as Naga Pidgin, are also clearly present (Sreedhar, 1974, 1975) and other varieties of non-European pidginized and creolized standard languages are also found, particularly those varieties of Hindustani described in Chatterji 1931 and Apte 1974. As we shall see, South Asianists working in this area also claim to have something to contribute to the theory of the genesis of creoles and pidgins.

A more sophisticated approach to the classification of language in general is exemplified by the article by William Stewart already mentioned (Stewart, 1962b), where he provides a typology based on not only the type of language but its function. His typology is important because of his recognition of language 'attitudes' that are involved in the formation of it, e.g. whether a language exhibits historicity, is standardized, has vitality (living speakers), and homogenicity (its vocabulary and grammar come from the same historical source). As we have seen in the discussion of diglossia, some of these attitudes underlie policies toward various forms of speech in South Asia and the conflicts associated with them.

Another article with some bearing on South Asia is that by Robson (1975) in which she introduces the notion of 'effability' of pidgins, i.e. a stage that a pidgin reaches at which point it is capable of expressing everything that its users want to say. Robson claims that if pidgins attain effability, they are then no different from natural languages (non-pidgins, non-creoles), and that pidgins which never become effable never become creoles. There may be circularity in this statement, (any pidgin that doesn't make it into a creole is automatically ineffable) but given that fact that Naga Pidgin, for example, seems to be coming very close to a state of effability, and may be creolized within the next generation, some notion of 'viability' of a pidgin/creole needs to be developed.

Another theoretical article based mainly on data from Portuguese Creoles, including most Indo-Portuguese Creoles, is that of Laurenţiu Theban (1975) in which he attempts to abstract some universal semantic constructs from Creole syntax. This article does not have much to tell us about sociolinguistic usage of Indo-Portuguese Creoles in the subcontinent, but it does have some interesting data, especially on the loss of the ergative system found in Hindi (and Marathi, Gujarati and Konkani) but lost in Bombay and Calcutta Hindustani Pidgins and Fijian Hindustani. Theban claims that the reason that most of these Creoles show great similarities is not because of preservation of various features from their original donor languages, but because of universals of semantics that emerge when language systems are pidginized and creolized. Since many of his examples are from Indo-Portuguese, his article may have something to contribute to the theory of Creolization and Pidginization.

An important article for this subject as related to South Asia is that of Nida and Fehderau (1970) in which they introduce the notion of a '*koiné*'. They show that many languages have a somewhat reduced form that is commonly in use around the area in contact with the main (standard) language. They call this reduced form a *koiné* after the Greek *koiné* that was widely used in the eastern Mediterranean in the early Christian era. While this is not a pidgin, it is not exactly the same as the main or classical language, either. It is mutually intelligible with the standard, and has some social prestige, whereas pidgins are never mutually intelligible with the standard (aggressor) language and are of course very reduced with very little prestige. Although their examples are mainly from Swahili and other African languages, it seems to us that the variety of English in use in South Asia is a kind of *koiné* of English, since it is usually (at least standard South Asian English) mutually intelligible with British and American varieties of English (Bansal 1969) but still exhibits reduction of various sorts. Certainly pidginized varieties of English can also be found in the subcontinent, but it may be that there is a kind of continuum of speech forms, ranging from true pidgin to koiné English and in a few cases, to standard R.P. It seems to us that since this notion has been overlooked by those working on the subject of Indian and South Asian English, it is time for some mention of this to be made.

6.4.2. Creolization, Pidginization and South Asian Languages in General

There are a number of works on the subject of pidginization and creolization of languages in South Asia that make general claims as well as describing some situation involving this phenomenon. Gumperz (1964a) and Gumperz and Wilson (1971) have addressed the question of what kinds of differences can really be said to exist between pidgins and other (homogenetic) languages. As Gumperz (1964a) says it may be the case that structural borrowings occur very widely between languages but are never reported in the grammars because of the "existence of social norms" mitigating against the reporting of these borrowings in the grammars, whereas in the pidgin situation the high incidence of structural borrowing may be due to absence of these cultural norms, either against borrowing or against reporting the borrowing. One has only to review the controversy surrounding the publication of Hesseling's *Het Afrikaans* (1899) in which he proposed that Afrikaans was a relexified Portuguese creole, (a notion which was greeted with outrage in South Africa) to realize what political emotions may be aroused by proposing that some language has humbler origins in a pidgin or creole. In South Asia we have the example of Gandhi's proposal that Hindustani, rather than Hindi or Urdu, should be the national language of independent India, which has been followed by the relexification of Hindustani using Sanskrit vocabulary, a process that has made Sanskritized Hindi unintelligible to speakers of Khari Boli, to say nothing of speakers of Hindustani. Yet given the South Asian predeliction for diglossia, especially diglossia with archaic vocabulary, it is not hard to see why this had to happen. Hindustani simply lacked the prestige (and in the eyes of many speakers of other languages, even Sanskritized Hindi still lacks the prestige of e.g. Tamil, because of a lack of an ancient literature).[26]

In the article by Gumperz and Wilson, which examines the convergence of Kannada, Marathi, and Urdu in Sangli District Maharashtra, the authors are led to conclude that since there is no evidence of pidginization having occurred, but only convergence of the three systems over the centuries, we seem to get a kind of creolization without pidginization.[27] Thus the authors are able to say that "there seems to be no reason therefore to draw an a priori distinction among pidginization, creolization and other diffusion processes; the difference may be merely one of degree."[28] This paper is important because of its recognition that creolization, at least in many South Asian contexts, is related to bilingualism.

Another very important study of pidginization and creolization in South Asia is that of Southworth (1971); important because of its claim that Marathi may have originated as a pidgin, and especially because Marathi is a standard language recognized by the Indian Constitution, enjoys prestige as the official language of Maharashtra, and has extensive literature. We do not know what kind of reception Southworth's thesis may have received, or may yet receive, in official circles in Bombay and Poona, but we doubt that the reception will be warm.

> Marathi, even in its oldest known form (tenth century A.D.) . . . presents . . . a picture [of extensive non-lexical resemblance to other languages]: grammatical and semantic resemblances with Dravidian are massive, but there are few actual lexical items from Dravidian sources. The vague term 'influence' is often used to describe such phenomena; Boas's phrase, 'the diffusion of grammatical processes over contiguous areas' (1929:6) is equally unhelpful in contributing to our understanding of the social processes involved. Clearly, not all cases of influence imply pidginization, and as this question is pursued more deeply it becomes clear that we need a more precise sociolinguistic typology of outcomes of language contact. (Southworth, 1971:256)

Southworth also mentions how diglossia is involved in complicating the picture of whether Marathi is homogenetic or heterogenetic. Since "all writing and scholarship has traditionally been in the hands of Brahmans, the guardians of the purist tradition. . ." we do not find pidginized forms of Marathi or other Indo-Aryan languages in earlier texts. But today, "the upper class, while fiercely maintaining the purity of their ritual language, can often afford to take a much more relaxed attitude about their language of worldly intercourse, which has no religious or intellectual significance. . . . Thus we can expect to find the upper class much more tolerant of modifications in Prakrit introduced by others, and less motivated to preserve its purity; this in fact, coincides with the attitudes of modern educated Indians."[29]

Thus we see that at one end of the spectrum of caste dialects of Marathi we could have pidginization, while at the other end of the spectrum of dialects we have rigid adherence to the prestige forms of the language, such that one end of the continuum resembles Dravidian very strongly, while the other is clearly Indo-Aryan. Modifications percolate up from the Dravidian element, and are tolerated in L forms of Marathi spoken by Marathis in L contexts; only very slowly and imperceptibly do they never creep into H forms of Marathi.[30]

Southworth's detailed examination of the extent of Dravidian structural borrowings in Marathi can not be recapitulated here, but much of what he shows can also be postulated to have occurred in other Indo-Aryan languages in their contact with speakers of what were probably Dravidian languages and others as the Indo-Aryans gradually extended their influence over much of north India. The widespread structural similarities between all the languages of South Asia is reviewed in our chapter on the South Asian Linguistic Area (Chapter IV); what is important in Southworth's article is that the Dravidianization of Marathi seems to be much more extensive than any of the other Indo-Aryan languages, probably due to its further extension into South India and the Deccan than any other.

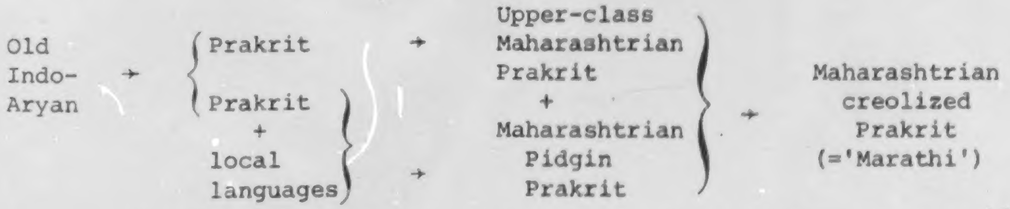
Southworth hedges in his final dictum as to whether Marathi is to be considered a true creole or not, but feels that "whether or not Marathi qualifies as a true creole, its present characteristics are probably the result of a prolonged process of mutual adaption between an Aryan language and a local pidgin-creole (or more likely a series of pidgin-creoles)."[31] His estimation of how this process developed bears repeating:

> As the Aryan language spread into the indigenous (Dravidian and/or Munda) speech communities, pidginized forms of Indo-Aryan were created and stabilized in the context of trade and joint agricultural activities within multi-caste settlements. The creation of these pidgins may have in part coincided with the consolidation of these settlements. In the initial stages the languages of such communities would have been: Sanskrit (confined mainly to ritual activities), Prakrit (i.e. a colloquial form of early Indo-Aryan), local language(s), and pidgin. The pidgin would have been the principal medium of communication between poor or middle-class cultivators and their superiors. . . .[32]

Then, as local people gave up the local languages for Prakrit or pidgin, "There was a gradual convergence between pidgin and Prakrit, to the point where they were no longer sharply distinguishable from each other, but were simply the extreme points of a continuum."[33] David DeCamp has posited the notion of 'post-creole speech continuum' (in Hymes, 1971) for Jamaica, and Southworth concludes that "Marathi may well be an example of a post-creole speech continuum of much earlier origin than any discussed hitherto."[34] "The result of this process would of course vary depending on the relative proportion of speakers of the different languages in each area; the same process was presumably taking place all over the present Indo-Aryan region, but with different degrees of pidginization in each place. Marathi (and particularly the lower-caste varieties of Marathi in the extreme south) apparently represents the most extreme, or most highly pidginized, of these developments."[35]

It is instructive to recall that the variety of Marathi examined by Gumperz and Wilson *supra* is an example of one of these extreme cases, with convergence of the local variety on a total scale with the local variety of Kannada, except for lexical items.

Southworth's schematization of this is as follows:


Old Indo-Aryan
Prakrit
Prakrit + local languages
Upper-class Maharashtrian Prakrit
+
Maharashtrian Pidgin Prakrit
Maharashtrian creolized Prakrit (='Marathi')

Southworth specifically rejects the kind of "bilingualism, involving fuller control of two languages on the part of a substantial segment of the population,"[36] that Emeneau (1962a) posits for Brahui and Balochi, because of the different results apparently produced by the two different situations. Also, because Marathi represents an extreme tendency of which Hindi is also a lower-degree example, it would be hard to distinguish between pidginization in Marathi and borrowing in Hindi, since elsewhere the two processes are believed to be separate. Southworth concludes that "pidginization took place throughout the Indo-Aryan area, but. . .its long-range linguistic effects were tempered or reinforced by other social factors (caste structure, diglossia, and Sanskritization); these factors have led, at the extreme end of the spectrum, to a result which is similar to the classic modern cases of pidginization known from the Caribbean and the Pacific."[37] Thus we see that in order to explain the enormous complexity of the variation in Marathi social and geographical dialects, one researcher has posited a pidginization process complicated by and tempered with other processes well-known in the South Asian area, such as diglossia. The latter seems to act as a vector moving against pidginization, or at least balancing it; indeed in Jamaica (DeCamp, 1971) the presence of standard English has obviously also led to the emergence of the continuum. The importance of Southworth's claim is that pidginization has been operative in the meeting of Dravidian and Indo-Aryan languages over much of North India (and in Sri Lanka also, we might add; cf. Elizarenkova 1972) and that the structural borrowing and convergence in the South Asian area is due in large part to this process. This is a strong claim, and much more work needs to be done to substantiate it. It is not a popular notion to subscribe to, but it may well be a fertile area to investigate.

6.4.3. Creolization and Pidginization and Specific South Asian Languages

Much of the data from the previous article really deserves to be treated in this section, because along with the theoretical claims there is a large body of data illustrating the extensiveness of pidginization in Marathi. Rather than devote any more space to that very accessible data, we restrict ourselves here to discussions of some other examples of pidginization and creolization in both India and Sri Lanka.

6.4.3.1. Indo-Portuguese

The earliest study of Indo-Portuguese is that of the early Creolist Schuchardt (Allgemeineres. . .Indo-Portugiesische) in 1889. His account is primarily historical and descriptive, with examples of some texts, estimates of number of speakers in various localities, and estimates of previous colonies of speakers no longer extant. He classifies

Indo-Portuguese into four types: 1. Gauroportuguese; 2. Dravidoportuguese ("diese beiden pflegte man bisher unter "Indoportugiesisch" zu verstehen"); 3. Malayoportuguese; and 4. Sino-Portuguese. Of interest to to us, of course, are the first two types. Gauroportuguese seems to refer to those varieties of Portuguese Creole spoken in primarily Indo-Aryan speech areas; Dravidoportuguese of course is spoken in primarily Dravidian areas, including Sri Lanka, since it is strongest there in Tamil speaking areas, e.g. Batticaloa, Jaffna, Trincomalee, Negombo, Mannaar (and of course Colombo, Galle, and Kalutara in the Sinhala area). Schuchardt recognizes that in many cases different kinds of Portuguese would have been spoken in the same locality; a standard Portuguese indistinguishable from that spoken in Lisbon; a 'halpverdorbene' kind, and a 'ganzverdorbene' variety, corresponding respectively to the speech of Portuguese and their offspring, the second to people of mixed blood, and the third (completely broken) by indigenous speakers of other languages. The third variety, however, was also used by other colonials (English, Dutch, Danes, French, etc.) in dealing with local traders, so that it was long the lingua franca was used by Europeans in contact with indigenous peoples. While Schuchardt seems to be somewhat of the notion that language and skin color or blood type have some connection (he devotes a great deal of space to such a discussions even though he eventually denies any connection[38]), he does eventually give some description of the varieties of Indo-Portuegese found in the subcontinent. He is also convinced that Portuguese Creole only persists where Portuguese is used as a standard language, and when no schools, sermons, catechisms or whatever are using Portuguese any longer, the Creole begins to be replaced by local languages. The importance of the reinforcement of the spoken language by the written may be salient; but this does not explain why Creole very early died out in Goa, the largest in area of the Portuguese enclaves in South Asia. Schuchardt attributes this to a policy of forced Portuguesization of Goa, in the notorious order of 1684, whereby the Viceroy Count de Alvor tried to stamp out the local language by ordering that the local population had to have adopted Portuguese within three years or suffer harsh penalties. Nothing resulted from this order, apparently, except possibly the harsh penalties.

More important for the gradual extinction of Indo-Portuguese in some localities was probably the protestantization of the previously converted populace by missionaries from other European countries. They apparently preached in Portuguese at first to woo the Creoles away from their Catholic parishes (which were weakened by little or no support from Portugal and elsewhere), then gradually in schools and in other uses English predominated, and the Portuguese population merged with English-speaking or other indigenous language groups.

Interestingly, recent reports by Theban[39] indicate that Portuguese Creole is alive and well in some areas where it was reported earlier to have ceased to exist. A description of Batticaloa Portuguese Creole is underway by Ian Smith[40]; apparently some speakers of Portuguese Creole have concealed the fact of its existence for various socio-linguistic reasons.

As far as linguistic description, Schuchardt does give many lexical forms and loan words from Indo-Portuguese into English, including some etymologies of items now found in Indian English, as well as some influences of English in Indo-Portuguese, e.g. *officina* for 'office' instead of *escriptorio* or *repartição*, as well as many out-and-out borrowings from English. Widespread bilingualism seems to have been the case for most "Portuguese" in South India, at least, since they needed to know indigenous languages to communicate with their neighbors: "So spechen die wenigen Portugiesen, die sich noch in Tranquebar befinden, Alle Tamulisch."[41] In the Portuguese period, Protestants and Catholics alike were of the notion that either 'good' Portuguese ought to be used in church and school, or not at all, since the Creole was held to be not only a 'lingua corrupta' but a 'lingua defectiva'. Apparently this attitude has caused any remnants of Indo-Portuguese to go underground except in areas where no attention is paid to it, because of the absolute lack of prestige of this one important lingua franca.

After Schuchardt came a number of studies by Portuguese scholars; early was Dalgado (1900), followed much later by Fonseca (1959). Dalgado noted the widespread bilingualism of the burghers of Ceylon, who 'all speak more than one' language[42] which has led to the importation of much 'exotica'. He also pointed out the resemblance of Ceylon Portuguese Creole with other Portuguese Creoles in India--the same archaisms preserved from earlier Portuguese, the same orientalisms which have been 'Portuguesized', the same indianized Portuguesisms, etc.

He noted the kinds of phonological changes from standard Portuguese to Creole, and attributed them to four causes: 1. indistinct pronunciation; 2. difficulty of articulation; 3. speed of enunciation; and 4. influence of foreign languages. But Dalgado was puzzled by the fact that one does not observe invariably the same application of these principles in an identical manner in all the dialects--"many Portuguese phonemes of the same nature assume different forms, sometimes without plausible reason."[43] Dalgado then gives a number of 'rules' which account for phonological changes[44]; under morphology, he notes the dative-stative use of the particle *tem*: *per mi tem hum livro* 'I have a book' (to me *tem* a book) instead of standard Portuguese *tenho um livro* which he says quite rightly is 'um reflexo das linguas indianas' which, including even Sanskrit, express by use of the verb *be* the idea of possession. Dalgado also mentions some varieties of Creole which form a 'bridge' between 'high' Portuguese and 'low' Portuguese, resulting in a miscellaneous hybrid language. Perhaps this is also an example of a Creole continuum described above by DeCamp (1971) and Southworth (1971).

The study by Fonseca (1959) is cursory and in the framework of a study of the history of the Portuguese language; only a few pages are devoted to Indo-Portuguese.

Dalgado also published a number of studies of individual dialects of Indo-Portuguese such as 'Dialecto Indo-Português do Norte' (1906) and 'Dialecto Indo Português de Negapatão' (1917). In the former he notes that the term 'Gauro-Portuguese' employed by Schuchardt is properly '*Gauda*'[45] and is very old in India, having relation to classes of Brahmans, who were divided into *pancha-gauda* and *pancha-dravida*, that is, five northern branches and five southern branches. The distinction was purely geographical, because the southern group also includes two 'Aryan' regions: Maharastra and Gujarat. The terms, Dalgado claims, are now used primarily by Europeans,[46] the Gaurian for Indo-Aryan, the Dravidian for what Dalgado calls Turanian.[47] The rest of Dalgado's studies are primarily descriptive and textual, although interest for the comparativist and the Creolist.

Indo-Portuguese is of interest recently to Creolists bent on proving one or another theory of Creole genesis. Thompson (1961) has done a classical 'monogenesis' study of old world Portuguese Creoles, capitalizing on Whinnom's *Spanish Contact Vernaculars in the Philippine Islands (1956)* and mentioning some aspects of the Macanese dialect of Portuguese Creole spoken now only in Hong Kong (as Goa, the indigenous Creole is spoken now only outside the original area, since, as we have mentioned earlier, Po(rtuguese) Creole in Asia seems to flourish only when no attention is paid to it). He points out many similarities in the Po Creoles of the Old World, claiming for instance that Po Creole comes to Macao fully developed and showing "little evidence of Cantonese influences in the structure and lexicon of their [i.e. Cantonese] mother tongue which seems to have come to China ready-made. Its structural similarities to the Malayo-Portuguese dialects of Malacca and Java, to the Indo-Portuguese complex and the Portuguese creoles of West Africa. . .are much more numerous than the occasional resemblances to Cantonese structure."[48] He particularly remarks about the particles used for aspect markers and how they are similar in Po Creole, and Jamaican, Haitian, Dominican, Saramaccan, etc. Creoles. Thompson would find it "exciting" if it could be proven that a universal Creole Grammer, that one day might be set up, "was a development of a Mediterranean lingua franca."[49]

Thus the devotion to the thoery of monogenesis leads Thompson, and as we shall see, others, to ignore local influences and to see only the shining examples which prove his point.

Lexical studies, often concerned primarily with dating borrowings, are typified by Knowlton's "Portuguese-Tamil Linguistic Contacts" (1969). He devotes space to both borrowings from Portuguese into Tamil and from Tamil, Malayalam and Sinhala into Portuguese. As might be expected, the borrowings fall into categories of primarily culture items found in the donor language but not in the receiving language. He also devotes some space to discussing how items may be borrowed through intermediaries, e.g. Tamil to Portuguese to English, or Portuguese through Tamil to Malay.

The most up-to-date studies of Indo-Portuguese have been done by Maria and Laurenţiu Theban; only M. Theban's "Structura Propoziţiei în Portugheză şi Indo-Portugheză" (1973) is available in print, although a later paper was read at the Pidgins and Creoles conference in Honolulu in 1975 which we were able to hear. The Thebans seem to be interested in proving some generalizations about universals in Creolization, so they tend, like Thompson before them, to observe only similarities between Portuguese Creoles, and no similarities between Indo-Portuguese and indigenous languages; this despite observations (L. Theban, 1975) that there is a dative construction involving the particle *tem* similar to the 'dative-stative' constructions in Dravidian and other languages (Schiffman, 1970a) of the subcontinent where stative verbs take a dative subject ('to-me it is liked/available/understood/known etc.').

Ian Smith (1975) has indicated that while variation in the syntax of the dialect he examined[50] does exist, there are clear examples of the influence of Tamil in dative-stative constructions other than with the copula *tem*. For example with the verb *intinda*, 'understand' a dative seems to be required: *portugēs etus tudus pa lo intinda* (Portuguese
1 2 3 4 5 6 1
they all dative future understand) 'They all understand Portuguese.'
2 3 4 5 6

With the verbs *nistāy* 'want, need' and *kera* 'want', dative constructions are also observed: *nōs pa nistāy diyānti su pɔ̄nta namās mē nā* (we dative
1 2 3 4 5 6 7 8 9 1 2
want front genitive end only emphatic no) 'We want the front end only,
3 4 5 6 7 8 9
don't we.' *ēli pa ōtru kī kera* (he dative other what want) 'What else
1 2 3 4 4 1 2 3 4 5
does he want?' Other examples involving the verb *ačā* 'get', and *suwā* 'sweat', also have dative subjects *parim* (*parmi*) 'to me'. In fact, M. Theban even gives an example herself of a dative construction involving the verb *sabē* 'know' (1973:642): *Etros-pa minh condição sabē* (to-them my condition is known) 'they know my condition'.

It seems to us, therefore, that rather than making sweeping generalizations about how there are no grammatical formalities that are non-Romance (M. Theban, 1975) in Indo-Portuguese, there should be some serious comparative work done on the dialects still extant. If the Creole theory still holds that a Creole language is typically the structure of one language with the vocabulary of another, and if the Theban's thesis that Indo-Portuguese is both Romance in deep structure and in lexicon, why is Indo-Portuguese not just another mutually intelligible dialect of Portuguese (which it is not) instead of a separate mutually unintelligible language (which it seems to be)?

6.4.3.2. Studies of Pidginized Hindustani

Another South Asian speech form which has long been recognized to be a reduced version of something else is Hindustani. One of the earliest serious studies of this pidginized version of Hindi is that of S. K. Chatterji (1931). He begins by treating the history of the rise of Hindustani/Hindi/Urdu and its spread to Calcutta, where it was the native language of some, but not of many, of the first inhabitants of early Calcutta; Bengalis and others had to learn some varieties for communication with government representatives. In other words, the early history of Calcutta Hindustani has to do with the development of Calcutta and the relationship of rulers to ruled there. Nowadays a reduced variety of this earlier form, called Bazaar Hindustani, exists in Calcutta and acts as the lingua franca of all *non*-Bengali groups (except Oriya speakers) there, and even some Hindi speakers use it with non-Hindi speakers. Chatterji illustrates this by giving examples of different caste and occupational groups and what they speak with whom in Calcutta.

Chatterji posits a very mixed history for Calcutta Hindustani: it arose when "peoples of North India speaking at home Lahndī, Punjabī, Rājasthānī, Brajbhākha, Kanaujī, Bundēlī, Awadhī, Bhōjpuriyā, and Magahī had already found Hindustani a common *lingua franca* which they all gladly recognized."[51] "The 'Jargon Hindustani' as used (without the least regard for Delhi usage) by a Bihārī or a Purabiyā or a Mārwaṛī, was bad enough, and in Bengali mouths it took a further colouring from the Bengali Speech. . . .A certain attempt at accomodation with the Bengali language also unconsciously affected the Bāzār Hindustani of the Up-country people to whom the accent and words and forms of Bengali were becoming more familiar;"[52] thus a Bengali norm of B(azaar) H(industani) became established, as a compromise between Bengalis' attempt at Hindustani and upcountry attempts to adapt their bad Hindustani to Bengalis' attempt to understand. "Calcutta may be described as a bilingual city, Bengali and Hindustani being its predominant native languages." Apparently foreigners in Calcutta usually learn BH, rather than Bengali, while some Indian groups learn Bengali as well as BH. Chatterji describes BH of Calcutta as having been "living largely on a background of Bengali. At times it may be described as just a compromise language between Bengali and Hindustani. . . .already coloured by Eastern Hindi and Bihārī, and further modified in Bengal under the influence of Bengali vocabulary and idiom."[53] Eastern Hindi and Bihari elements are found mainly in the morphology and vocabulary; Bengali influence is primarily in vocabulary.

"A simplified Eastern Standard of Hindustani in fact may be said to be in existence. In it, grammatical gender is ignored; and the passive and neuter constructions of the transitive verb in the past tense, which is so characteristic of Western Hindi, have been done away with."[54] Chatterji describes other phonological, morphological, and lexical changes in great detail and also gives texts.

An interesting comparison of Bombay Hindustani and Calcutta Hindustani can be made by looking at Chernyshev's article on the former (1971) in which he touches on a few sociolinguistic aspects of Bombay Hindustani, in what is essentially a classical 'description-of-the-dialect'. He compares Bombay Hindustani with standard khaṛī bolī and occasionally with Calcutta Hindustani, based on Chatterji's work. He attributed many of the differences in Bombay Hindustani to the effect of bilingualism in Maharashtra and Bombay, and how it is that whenever people are bilingual there, the second language is almost always Hindi. So Hindi in Bombay is almost always spoken by bilinguals, since there are very few speakers of standard khaṛī bolī there. Therefore Bombay Hindustani is highly influenced by other languages, is reduced in various ways, such as lacking aspiration of consonants, elision of /h/ intervocalically, loss of case, number, person, and gender in nouns and verbs; changes of nasalized /ẽ/ to /i/ in the future; elision of /h/ in /hai/ in the present-future (/hai/ → /ay/); and loss of ergative constructions in syntax, except for some use by 'educated' people. Since there is invariability of nouns and verbs, syntax is very dependent on the use of postpositions and word order. He devotes much attention in fact to the postpositions, since they seem to be all-important in syntax.

An interesting aspect of this study is his recognition that there is variability in the structure and use of Bombay Hindustani depending on the educational level of the speakers, e.g. the use of ergative construction by better-educated people (presumably who have already studied some standard Hindi), but mostly the influence of Marathi is very strong, and some of these Marathi influences and others are percolating their way up into the standard Hindi used in Bombay alongside Bombay Hindustani. Thus there is developing a Bombay dialect of Hindi (as well as a Bombay variety of Hindustani that shows similarities with Calcutta Hindustani) but with a Marathi coloring instead of a Bengali coloring.

Another study of Bombay Hindi-Urdu written at about the same time as Chernyshev's but published later is that of Apte (1974). Like Chernyshev, he distinguishes two levels of usage in Bombay, one 'Level I' type spoken by educated people, who make an effort to speak standard Hindi or Urdu, and 'Level II' type, spoken by uneducated people who have little control of grammatical proficiency. Apte describes the features of Level II Bombay Hindi-Urdu, (he eschews the use of the term 'Hindustani') and compares it with Level I Hindi-Urdu. He rightly notes that many of the features of pidginization are present here, including a 'continuum' of lects from pidgin all the way to standard Hindi-Urdu, like that described for Jamaica by DeCamp (1971). Noteworthy is Apte's concern for some theoretical issues, such as:

> (1) are there any special linguistic and sociolinguistic features of Bombay Hindi-Urdu which sets it apart from other pidgin or creole languages, and what plausible explanations can be provided for these features; (2) Should Bombay Hindi-Urdu be considered a process or a stable final product of pidginization; and (3) what descriptive label can be used for Bombay Hindi-Urdu and how can it be classified.[55]

In contemplation of the first question, Apte points out that it is not just in Bombay and Calcutta Hindi-Urdu that reduction takes place, but also in some other areas (supposedly not pidginized dialects) such as in Sangli district (Gumperz and Wilson 1971, cf. Chapter VI. A) and in urban lects of Hindi-Urdu within Hindi States. Secondly it is difficult to say at this stage what 'standard' Hindi-Urdu is because of the many different varieties, from Sanskritized to Persianized down to the variety under study here, that exist. Thirdly it is not yet possible to say when a language has been simplified enough to be called a true pidgin. Futhermore, it is also not possible to state at this point whether Bombay Hindi-Urdu is a pidgin or is just in the process of being pidginized, i.e. is it a process or a stage? As Apte puts it,

> only the most exhaustive analysis of extensive material will reveal if the Bombay Hindi-Urdu speech described here can be named with any such labels a convergence, pre-pidgin continuum (Hymes, 1971:68), salient pidginization, or substantive pidginization (Samarin, 1971:119).
>
> If Bombay Hindi-Urdu is to be considered a stable end result of pidginization, we need to know the extent of its spread, the time-span of its existence, and perpetuation of its fixed structure in a consistent manner.[56]

Apte concludes with some recommendations for further study of Bombay Hindi-Urdu, and for a name for it: depending on whether BHU is still a process or has reached a stage, he recommends calling it *Bombay Pidginized Hindi-Urdu* or *Bombay Pidgin Hindi-Urdu*, respectively.

6.4.3.3. Naga Pidgin

Another pidgin that has been on the South Asian linguistic scene for some time, but has received little attention until recently is a pidginized form of Assamese spoken by Nagas in Nagaland, called variously Naga Pidgin, Pidgin Naga, Nagamese, and earlier, Nagaassamese. Its existence has been noted since the early 20's and 30's (Hutton, 1921; Haimendorf, 1936).[57] The only scholarly work on the subject seems to be that of M. V. Sreedhar (1974, 1975). Apparently Naga Pidgin arose because of the extreme linguistic diversity in the area, with many different Naga languages being spoken throughout Nagaland, such that even in areas where one or another major Naga language is spoken, many minor languages are also used. In other words, there are no monolingual areas in Nagaland; there are also no Naga languages with prestige enough to have been used as a standardlanguage. Since the nearest language with standardization and prestige was Assamese, it was the language which contributed its vocabulary as the basic lexicon for pidginized Naga. Nowadays the Bible has been translated into some of the Naga languages, e.g. the Ao are "both in terms of the written works produced and the percentage of literacy. . .in the forefront amongst the Nagas." (Sreedhar, 1974:18) Some of the other languages are used for primary education, but only Ao and Angami are allowed as mother tongues for the S.S.C. (Secondary School Certificate) Examination. Whenever language textbooks are not available in one minor language, other Naga languages may be used for the medium of instruction; but in many cases English is being used as early as Standard VI. This has developed into an interesting situation--because of inadequate preparation in English by both pupils and teachers, teachers often use Naga Pidgin as a link language to explain unofficially what they are teaching in English. But because Naga Pidgin lacks prestige, authorities are unwilling to use it as a medium of instruction, although it is filling this function unofficially. It is Sreedhar's opinion that Naga Pidgin should be used as widely as possible as a medium of instruction in Nagaland, since it is known by all Nagas from early childhood onward, and would enable children to learn through the medium that they are already unofficially learning through.

Sreedhar's 1974 work is concerned in part with a classification of the Naga languages; his classification differs somewhat from earlier classifications (e.g. Grierson), being based primarily on Marrison (1967). Sreedhar dwells on this aspect of language classification in order to clarify mainly the different varieties of Naga Pidgin which are in use, because the pidgin varies according to which MT speakers are using, and the resultant varieties tend to fall into groups along the MT lines of classification. Sreedhar begins by describing Naga Pidgin as spoken by Angami speakers; following this is a description of phonemic variants in the pidgin of different Naga communities. While all Nagas share the "four-way, three-way, and two-way oppositions in the place of articulation found respectively with stops, nasals, and approximants. . ." many communities do not share "voicing with the stops, opposition between the flap and lateral, and a three-way opposition in the place of articulation with the fricatives."[58] On the basis of these similarities and differences, Naga Pidgin may be grouped into three groups:

> i) the Southern group consisting of the speakers of Angami, Kachari [a non-Naga community residing in Nagaland], Zemi, Liangmei, Rengma, Rongmei, Sema, Chokri, and Mao, ii) the Northern group consisting of the speakers of Konyak, Sangtam, Phom, Chang and Khiamngans and iii) the Central group consisting of the speakers of Lotha, Ao, and Yimchunger.[59]

The Southern pidgin variety shows the greatest number of oppositions in manner of articulation, while the Northern Naga Pidgin shows the least number, with Central in between. It also develops that the Kachari community, a non-Naga group living in Nagaland, speaks a *creole* which differs somewhat from Southern Pidgin (SP). Sreedhar then inventories the minor differences found in the SP of Zemis, of Liangmeis, of Rongmeis, etc., the CP of various Central groups, and the differences in the NP of various Northern groups. In general, SP, CP, and NP correspond to Grierson's Western, Central, and Eastern groups, and to Marrison's types C, B, and A, respectively; but in one case, Sangtam is classified by Grierson as Central and Marrison as B, but it is actually a sub-member of Northern Pidgin on the basis of Sreedhar's comparison of phonological differences. Otherwise there is identity in the three classifications.

After describing the morphology of Naga Pidgin, Sreedhar discusses variation in some grammatical classes, especially variations in number and cases (gender does not vary). For example (as with many pidgins and creoles the world over), in Naga Pidgin, when a noun is quantified with a number or quantifier, plural need not be marked. This is true of all varieties except Yimchunger Pidgin, which does mark the plural even when quantified, e.g. /suali/ 'girl' /sualik^{h}an/ 'girls' /bisi sualik^{h}an/ 'many girls' /sualik^{h}an duyta/ 'two girls'. In some varieties, the plural marker is absent with certain categories of nouns even when unquantified, e.g. "with the nouns referring to birds in the Ao, Chokri, and the Phom varieties.[60]

thus:	/guru/	'cow'	/gurukhan/	'cows'
but:	/suray/	'bird, birds'		
	/ciriya/	'sparrow, sparrows'		

Much more variation with plural markers and their allomorphs /khan/ and /bilək/ can be found in other varieties. The latter is, incidentally, the plural marker in Western Assamese, but the source of the former, /khan/, is unknown.

Sreedhar also notes variation in the use of case markers, in the use of negative particles, in the pattern of interrogation, especially in tag questions, in the form of the tense marker in the copula, in aspectual oppositions, and in the number and marking of modals in the various types of Naga Pidgin. He concludes his description by providing tables showing various differences in phonology, variation in case morphemes, and the other variables mentioned above.

This is a valuable work on a hitherto undescribed pidgin found only in the South Asian area; the title, A Sociolinguistic Study, leaves us somewhat disappointed, however, since there is very little sociolinguistics in the work other than the listing of variation. Hopefully Sreedhar or others will find an opportunity to study the actual dynamics of language use in Nagaland, for example, situations involving switching from mother tongue to pidgin and back; we look forward to further reports on the creolization of Naga Pidgin by the Kacharis and others.

Sreedhar's other work on this subject (1975), discusses variation of the sort mentioned above, with a view toward making some recommendations as to which variety of Naga Pidgin should be chosen as standard, in order to begin using it as a medium of instruction in schools. His recommendation is to use Southern Naga Pidgin since it has the most oppositions in manner or articulation, which would make it easier for Nagas to learn Hindi and English later, and because it is spoken in and around the State capital. His recommendations are couched in a manner which indicates he does not think that they are likely to be accepted, i.e., Naga Pidgin is not about to be used in any variety as an official language of Nagaland, even though it probably makes greater sense to use it than any other language. Some method obviously needs to be found to invest it with more prestige in order to help with its acceptance.

6.4.3.4. Vedda Creole in Sri Lanka

We are fortunate in having a very recent analysis of the speech of the Veddas in Sri Lanka by Dharmadasa (1974). He reviews the work of Nevill and Marrambe who felt that the speech of the Veddas was probably a language distant from Sinhala, whereas later scholars such as Parker, the Seligmanns and Geiger felt it was a dialect of Sinhalese.[60] The notion that what the Veddas speak was neither a separate language nor a dialect of Sinhala was advanced by Sugathapala De Silva, i.e., that it was a creolized language arising out of contact between the Vedda's original language (now lost) and Sinhala. This theory helps account for a number of aspects of Vedda which distinguish it from both dialect and separate language status. Dharmadasa follows Sugathapala De Silva in recognizing the creole theory, and proceeds to show in what ways Vedda speech displays creole-like features. These are found in phonology, morphology, syntax, and the lexicon of Vedda. While the phonemic inventory of Vedda is similar to that of colloquial Sinhala, phonetic realizations of some sounds are of a different frequency than in Sinhala, e.g. Vedda prefers the palatal affricates /c/ and /j/, and especially replaces Sinhala /s/ with /c/. Dharmadasa feels that in the original substrate Vedda language, the palatals must have been of higher frequency than they were in the donor language (which he shows to be probably middle Sinhala).

Like other contact languages, Vedda shows morphological suffixes on nouns which act as classifiers, similar perhaps to /fela/ in some of the English pidgins and creoles. These suffixes are derived from individual Sinhala lexical items as well as from Sinhala nominal suffixes, for which Dharmadasa provides probable etymologies, e.g. /pojja/ from Si. /poḍḍa/ 'a little', /gejja/ from Si. /geḍiya/ 'nut'. Vedda Creole also shows reduction and simplification when compared with Sinhala: for example Si. /gonaa/ 'bull', /eledena/ 'cow', but Vedda /gonaa/ 'bull', /gonii/ 'cow'. Unlike Si., Vedda simply changes the final suffix vowel, while Si. either has suppletive forms, or introduces vowel harmony as well, e.g. Si. /uura/ 'pig', /iiri/ 'sow' Vedda and /uura/ 'pig', /uuri/ 'sow'. Vedda also seems to be coalescing the dative and locative, i.e. transferring some functions to dative which are locative in Si. Also, Vedda uses dative cases where Si. uses postpositions such as /gäna/ and /nisaa/. Vedda also lacks a distinction in numerals involving the categories animate and inanimate (nominal and adverbial) which Si. has, so Vedda manages with one numeral where Si. uses three; Vedda also loses distinctions in pronouns which Si. has four grades of. Like many pidgins, Vedda has periphrastic constructions where Si. has separate lexical items, e.g. Si. /vässa/ 'rain', Vedda /uḍatanin mandovena diyaraaccaa/ 'water falling from above', Si. /pänsala/ 'pencil' Vedda /karukurugaccana ulpojja/ 'spike making the sound *karu kuru*'. Thus, "although Sinhalese served as the source language in the formation of Vedda Creole, the latter has frequently preferred to coin new expressions of a descriptive character out of the already existing lexical stock rather than borrow the relevant word from the former."[61] Apparently contact with Si. has been arrested, so that further loans are not flowing toward Vedda. Furthermore the Veddas seem to have developed a resistance to Sinhalese and to acculturation in general, so that further borrowings have been even more inhibited. This again points out the difference between South Asian groups and others that have been mentioned, and gives validity to Emeneau's argument for different dynamics regarding the borrowing from prestige languages in South Asia (Emeneau, 1954).

Dharmadasa concludes by presenting evidence for the time of the formation of Vedda Creole as being probably in the period between the tenth and 16th centuries, since it displays archaic features of Si. known to have existed only during those periods. This corresponds to the migration of the Si. population from the dry zone to the southwest after the collapse of the dry zone civilization beginning around the 13th century and culminating in the establishment of the last Sinhalese kingdom in Kandy in the 16th century. After this period the contact was less since the Veddas seem to have moved east into the dry zone where they are now predominantly found.

Dharmadasa also gives reasons for the maintenance of Vedda Creole rather than assimilation to Sinhala--while the Si. culture was more advanced, the Veddas valued their own culture and resisted assimilation. Furthermore, the Sinhalese seem to have had grudging respect for the

Veddas, even attributing high ancestry to them, although they were 'uncultured'. Thus both sides seem to have maintained social distance between each other and the creole has never been replaced by Sinhala.

6.4.3.5. Conclusion

We have examined five different kinds of creole/pidgin situations in South Asia; Marathi, arising out of contact between Dravidian and Prakrit(s); Indo-Portuguese, involving contact between a European language and a number of South Asian languages; Hindustani, involving contact between Hindi-Urdu and other Indo-Aryan languages; Naga Pidgin, involving contact between a number of Tibeto-Burman languages and Indo-Aryan Assamese; and Vedda Creole, involving contact between Vedda, whose ancestry is unknown, and Sinhala, an Indo-Aryan language. A number of interesting problems arise with regard to these speech forms. One is the difficulty of knowing what status to accord the forms of Hindustani found outside the Hindi-Urdu area, since they show features reminiscent of recognized dialects of e.g. Marathi, as spoken in border areas like Sangli District, Maharashtra. As Gumperz and Wilson observe, there is little difference that one can note between pidginization and creolization, on the one hand, and convergence of the Sangli type on the other. Weinreich has distinguished between contact which leads to convergence and contact which leads to pidginization and creolization (Weinreich, 1953:68-69), yet Southworth has shown that many different factors are involved, and the outcome may range over an entire spectrum of differences, all bearing the name 'Marathi' but displaying characteristics of different contact situations at each end. Clearly the idea of the 'Post-Creole continuum' (DeCamp, 1971) needs to be expanded to include other dimensions. We hope more work will be done in the near future in this area, and we hope Creolists will pay more attention, not to just the creoles and pidgins, but to the other languages of South Asia which have been involved in their genesis.

6.5. South Asian English and the Influence of English on South Asian Languages

The topic of English in South Asia (Indian English and Ceylon or Sri Lanka English) has been of interest to scholars for a number of decades. Perhaps the earliest scholarly attention paid to the subject is Yule and Burnell's *Hobson-Jobson, A Glossary of Anglo-Indian colloquial words and phrases* (1886, reprinted 1968). As they note in their introduction,

> Words of Indian origin have been insinuating themselves into English ever since the end of the reign of Elizabeth [I] and the beginning of that of King James, when such terms as calico, chintz, and gingham had already effected a lodgment in English warehouses and shops and were lying in wait for entrance into English literature. Such outlandish guests grew more frequent 120 years ago, when, soon after the middle of last century, the numbers of Englishmen in the Indian services, civil and military, expanded with the great acquisition of dominion then made by the company; and we meet them in vastly greater abundance than now.[62]

Their original intent was to catalogue words of Anglo-Indian origin in use in India, but this expanded over the years to include words of other origin, e.g. Portuguese, which had entered English by means of contact with India, and words of other origin which somehow found their way into English through an Indian language.

Valuable as this source is for the many etymologies and proposed sources for Anglo-Indian words in English, it is restricted in its usefulness as a picture of what English is like in South Asia, since it is at first and foremost a study of the *lexicon* of Anglo-Indianisms, and at that, Anglo-Indian in the older sense, i.e. English in the manner of people of European descent born or residing in India, rather than in today's sense, Indian of mixed European-Indian descent, and their mother tongue, Anglo-Indian English. We thus have already a two-way distinction, to which should be added a third: English as spoken by South Asians of non-European descent, for whom English is a second language, not a mother tongue. This third distinction has received the most attention in recent years, since the first category of speakers has dwindled away after Independence, and the second never did receive much serious attention, being considered somewhat lacking in prestige (Spencer, 1966).

6.5.1. Anglo-Indian English

The only article in the literature available to us on Anglo-Indian English in the modern sense has been that of Spencer (1966). Its theoretical thrust is the 'interference' notion, that the speech of Anglo-Indians (formerly 'Eurasians') is due to the interference of the speech habits of the original mother tongues and 'father accents' of the ancestors of the present population of Anglo-Indians. That is, the mother tongues such as Bengali, and the 'father accents' such as Irish or Scots wrought their influence on Anglo-Indian.

Spencer notes that Anglo-Indian has never been very prestigious (it was derogatorily referred to as 'Chi-Chi (English)' by the British and others) and present-day Indian teachers of English think very poorly of the accent, having acquired this prejudice from the British. "This is evidence, if such is needed, of the capacity for socio-linguistic attitudes to be transferred in a situation of cultural contact, and for the social or aesthetic rationalisations of such attitudes, in terms of a framework of class and status which has no empirical validity for the borrowers, to be transferred along with them."[63]

It is noteworthy, from the point of view of variability in Anglo-Indian English, that Spencer finds very little regional variation. Apparently Anglo-Indians were very mobile and highly urbanized and tended to transfer from one Anglo-Indian school to another all over India; Spencer feels that the Anglo-Indian school (which was rarely attended by Europeans of any status) is responsible for also fixing the norm of Anglo-Indian English probably once and for all.

After a brief discussion of the salient features of Anglo-Indian English (such as a lack of aspiration of voiceless plosives in all positions, lack of retroflexion in /t, d, l, r/, monophthongization of mid vowel diphthongs /ei/ and /ou/, and centralization of diphthongs /ai/ and /au/) he notes that one of the most characteristic features of Anglo-Indian English is the prosodic differences in stress, pitch and syllable length which impressionistically add up to a distinct 'sing-song Anglo-Indian accent'. As to the origin of these features, Spencer lays them at the door of Bengal, since the Anglo-Indian community seems to have gotten its start there, and because the lack of retroflexion and the nature of prosodic features are reminiscent of Bengali to a strong degree.

This brings us to the bulk of studies of South Asian English (SAE), namely, studies of English as spoken by other tongue speakers of South Asian languages. There are a number of studies which concentrate on SAE as a unified system, and there are some studies of particular varieties of SAE such as Marathi English, etc. There is also an interesting work by Bansal, already mentioned earlier, on the "Intelligibility of Indian English" (1969).

Finally, there are a number of articles dealing with the effect of English borrowings on a number of South Asian languages. While this may seem like a completely different topic, we are faced with the dilemma of what to do with speech varieties involving so much code-switching that it may at times be difficult to decide which code one is dealing with. Southworth and Daswani give an example of this mixture, here of English and Tamil:

> Yes? aamaanko ('yes')-----*yaaru peecuratunko* ('Who is speaking?') -----I am. . .speaking, sir-----Good morning, sir-----Oh, yes----- Sure, sure-----*sari* ('all right')-----*enke saar niinka varratu* untimely-*yaa varriinko* ('How is it that you come at such untimely hours, sir?')-----You can always meet me in the office between eight and nine in the morning. . .*veere viseesam onnum illinkale?* ('nothing else special?') Thank you.[64]

6.5.2. *Indian English*

One of the scholars most interested in the study of SAE has been Braj Kachru (1966; 1969). The 1969 article summarizes aptly his main concerns with the subject. Kachru gives an overall picture of the situation of English in South Asia and examines where and how it is used; he claims that South Asian English is an abstraction of the same nature as Standard American and R. P. English. He raises the issue which has interested other writers, namely is SAE a 'dialect' of English like American and British dialects, or is it a pidginized version. As we have mentioned earlier, we prefer the notion of *koiné* (Nida and Fehderau, 1970) as most aptly capturing the role of standardized English in South Asia. Kachru does not use the term *dialect*, since that does not seem to be in use among British-trained liguists; he prefers the term *variety*, which he defines as follows:

> I shall use the term *variety* to mean two or more varieties of a language "developed" in different *contextual* settings. These may either be those varieties which are used as *first* or *primary* languages (e.g. American, English, British English, Canadian English), or those varieties which are used a second or *foreign* languages, (e.g. Indian English, Filipino English, West African English).[65]

Thus the question of whether SAE is on a par with e.g. American or British English is sidestepped by using the term *variety* for both first and second languages.

After tracing the history of the introduction of English into the subcontinent, Kachru approaches the question of diversity within the 'standardized' variety of SAE. Obviously there is a standard *variety* of SAE, used by educated people throughout South Asia, irrespective of their own mother tongues, i.e., this standard SAE would show little influence from different mother tongues in the area. Varieties of SAE which do show influence of different mother tongues would be classified as a regional sub-variety of SAE.

Kachru also introduces the notion of 'cline of bilingualism' in the studies of SAE--the notion that there is an inclined plane representing competence in English which ranges from zero to 100%; perhaps this is comparable to the continuum noted by DeCamp in Jamaica (Hymes, 1971), although the lower end of the cline would be different qualitatively and quantitatively from the lowest end of the continuum in Jamaica. Kachru notes that the variety used by educated speakers of SAE is at about the middle point of the cline--the large numbers of civil servants, teachers, etc.

Kachru gives a detailed summary of studies of 'South Asianness' at different levels-studies of South Asianness in phonology, morphology, lexis, style, and literature. Since his is already a summary of many articles etc. some of which are not valuable to us, we will not attempt to recapitulate here that work, but refer the reader to it as the best source for information on those aspects of SAE.

In his section "Sociolinguistics and South Asian English" Kachru touches upon a number of sociolinguistic ramifications of SAE, quoting from his own work on contextualization and collocation, defined as "The collocationally deviant formations are those formations which function in Indian English contextual units, and would perhaps be unintelligible to a native speaker of English only because he is not acquainted with Indian contexts of culture." (quoted from Kachru, 1966) Since the study of contextualization and collocation is mainly concerned with lexical expressions of system and structure in the British school of linguistics, Kachru does not devote any space to discussion of phonological expressions of sociolinguistic variation, since he feels that while such studies do exist, they are the result of the "undue emphasis"[66] that structural linguists paid to these topics. He defines "sociolinguistically significant aspects of SAE" as (1) South Asian registers of English and (2) South Asian speech functions. Kachru's own discussion and illustration of registers and functions unfortunately leaves us unclear as to their significance.

Finally, Kachru notes that "the English language has not only been *South Asianized* by the South Asian linguistic and sociological complexes, but, what is more important, on its part the English language has left a definite mark on the major S(outh) A(sian) L(anguages) and literatures,"[67] and concludes with an attempt to assess the future of English in South Asia.

In attempting to assess the impact of Kachru on the study of SAE, it is also difficult to make a judgement, since his work is couched in the terminology of the British School, and no explanations are ever offered the American-trained linguist as to how any insights might be interpreted in a framework that the reader is familiar with. Undoubtedly this objection could be answered with the response that the reader should acquaint himself with the theory, and that that is the responsibility of the writer. The subject of SAE therefore awaits study and interpretation by scholars familiar with all theories on the subject.

It seems to us that what is lacking so far in discussions of Indian English is a treatment of the social uses of English in India--code-switching, varieties of bilingualism, contextual uses of English vs. another mother tongue (in the Labovian sense, rather than the Firthian sense which Kachru uses) so that rather than positing the 'cline of bilingualism' which assumes that English is to be graded on a continuum of none to 100%, we would like to see a discussion of 'compartmentalization' of English usage, i.e. which social contexts English is used in, and when does a speaker switch to and from English.

6.5.3. Code Switching with English

Shanmugam Pillai has done a study of this (1974), which, as the title implies, concentrates on code-switching in Tamil literature, a process which probably reflects more or less what goes on in actual conversation. He delineates a number of different kinds of code-switching, such as personal CS, complete CS, sprinkled CS, educated sprinkled CS (in which different codes may be used in the same conversational slice), non-personal CS, and furthermore sprinkled code-switching may be reflected (a repetition of the same idea in the other code) or non-reflected. Illustrations of these terms are given in the study, one of which is reproduced here.[68]

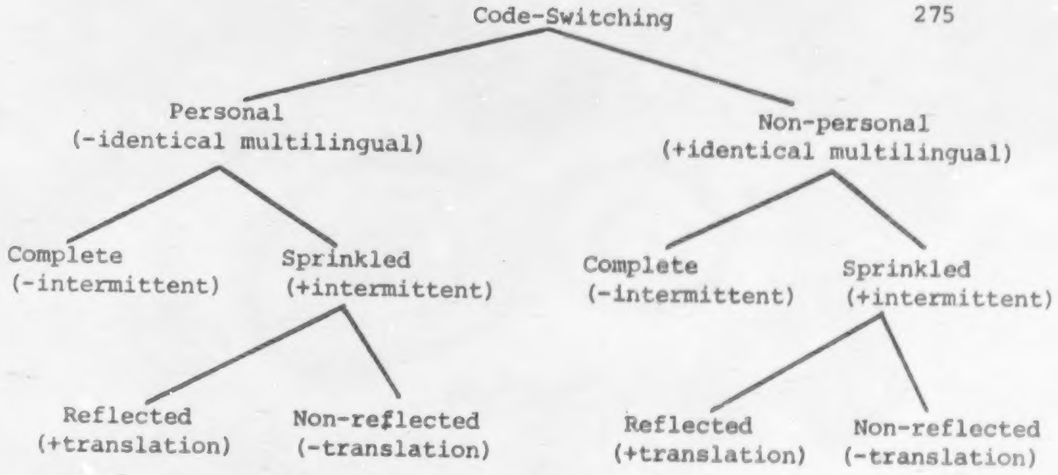

Code-Switching
Personal
(-identical multilingual)
Non-personal
(+identical multilingual)
Complete
(-intermittent)
Sprinkled
(+intermittent)
Complete
(-intermittent)
Sprinkled
(+intermittent)
Reflected
(+translation)
Non-reflected
(-translation)
Reflected
(+translation)
Non-reflected
(-translation)

Figure 60. Code switching in Tamil (Shanmugam Pillai, 1974).

In his taxonomy Shanmugam Pillai also presents a set of alternate distinctive features or components, e.g. personal code-switching could be given the feature "minus identical monolingual", etc. Shanmugam Pillai specified that the last criterion 'reflected' or 'non-reflected' code-switching is defined by whether the speaker is *conscious* of code-switching or not. If he is not, an English phrase in his Tamil, e.g. will be considered a borrowing; if he is conscious of code-switching, the English item is an instance of code-switching. Monolinguals will only exhibit borrowings, whereas bilinguals will exhibit both borrowings and code-switching.

Shanmugam Pillai's analysis provides a good basis for further study of code-switching in South Asian languages. He shows that code-switching may occur because of taboo items in Tamil, e.g. in the discussion of pregnancy and sexual intercourse; or because of 'intense emotion' on the part of one character or another; sprinkled code-switching is much more common on the telephone than in personal conversations, on the other hand. It remains to be seen what other factors one might be able to discover about this phenomenon which would lead to a serious and complete analysis of code-switching, especially involving English in the South Asian sociolinguistic scene.

6.5.4. Interference

Very common in the literature on SAE is the topic of interference in the English of South Asians from their mother tongues. Kachru (1969) cites a number of studies. We have examined a few that were available to us. Typical of these studies is Ashok Kelkar's "Marathi English: A

Study in Foreign Accent" (Kelkar, 1957). This is a classical American structural approach, a contrastive analysis showing how English structure and Marathi structure overlap and how some sort of compromise emerges. It has a useful appendix of amusing spelling pronunciations of English words by Marathi speakers. Another example is Pandit (1965), "Indian Readjustments in the English Consonant System," in which he gives the consonantal systems of both English and Indian languages and states how some consonants of English not found in South Asian languages are realized in those systems. One thing missing from many of these analyses is a study of phenomena not explainable by interference, e.g. in Tamil English there are replacements of English phonemes which cannot be explained by resorting to explanations of 'interference' since Tamil has some perfectly adequate phonological material which could substitute for the missing English sound. For example, the /oy/ diphthong in words like 'boy' etc., is regularly replaced by Tamil /aay/, so that /boy/ is pronounced something like [baːy] even though the sequence /oy/ and even /ooy/ occur in Tamil, e.g., /poy/ 'lie, falsehood', /pooy/ 'having gone'. There need to be studies of how it happens that English /oy/ is regularly replaced by Tamil /aay/, i.e., is it the case that because Tamils learn English from other Tamils who have in turn learned English from Tamil speakers, some distortion of the English system occurs over the generations, so that phonetic correspondences no longer have any weight in the matter? Or are there systematic phonological reasons which operate in borrowing situations, such that even if the phonetics of the two languages are similar, phonological constraints of the borrowing language distort the borrowed item? The latter seems to be the case in some instances, e.g. English 'glass' usually has the form /kLaas/ or /kiLaas/, with /k/ substituted for /g/, and /i/ inserted between the /k/ and /L/ (which incidentally is perceived as retroflex, possibly because of its 'dark' quality). But the Tamil phonological rule which lowers the high vowels when followed by /a/ in the next syllable (Bright, 1966a) also may operate on the sequence /kiLaas/ so that sometimes one hears /keLaas/ or even /keLas/ in the mouths of monolinguals for whom this item is a borrowing. It seems to us that this is an area which needs more study, i.e., the 'abstractness' of the level at which borrowing takes place, since interference from the surface phonetics of the mother tongue alone will not explain certain forms.

6.5.5. *Intelligibility Studies*

We have mentioned in passing a study by Bansal "The Intelligibility of Indian English" (1969) which is the first that we have noted that tries to measure the percentage of intelligibility of different kinds of South Asian English to native and non-native speakers of British and American English (and others). This study arose out of the concern for what kind of English should be taught to Indian teachers of English, since R. P. English is not felt by most Indians to be the kind of style they want to emulate, it being considered affected and stilted by most South Asians hearing it in the mouths of other South Asians. This concern was to develop a style of spoken English which would be comprehensible to American and British R. P. speakers, for example (and other non-native users of English) and would still be acceptable to Indians.

His conclusions are interesting and we will summarize a few of them here: R. P. and American speakers of English understand between 95% and 100% of connected speech where the context was known; it diminished considerably in unconnected speech where the context was not familiar to the hearer. Non-native speakers of English, *including other Indians* had *more* trouble understanding Indian English than did native speakers of English, especially when the Indian speakers were speakers of other mother tongues than the listeners. That is, Tamils for example have more trouble understanding Hindi English than do native-speakers of English. The average Indian listener rated only 74% comprehensibility for OT Indian English speech.

One of Bansal's most important findings is that the greatest problems in mutual intelligibility in Indian English stem, not from substitution of individual phonemes (although the substitution of stops for affricates (/t/ for /θ/) for example was a problem), but rather "The difference between the more intelligible and the less intelligible Indian speakers of English" lies "rather in the frequency of mistakes Indian speakers make in the distribution of vowels and consonants and in patterns of word stress, sentence stress, rhythm, and intonation."[69] Thus, an English word pronounced with proper stress but with faulty phonetics is more likely to be understood than a word pronounced with proper phonetics but faulty stress. The author concludes with some recommendations for improving the intelligibility of Indian English; for example, Indian speakers should pay attention to stress and rhythm, and should use a pronouncing dictionary. He also recommends more intelligibility testing for various dialects of English around the world. He does not come up with any recommendations for which variety of Indian English should become 'standard' Indian English, although that variety which was most easily understood by R. P. speakers was spoken by a Punjabi who had had native speakers of English (British) teachers throughout much of his education, and had spent much time listening to English on the radio and attending films, and therefore approximated very closely R. P. speech. More work obviously needs to be done in this area before a 'standard' Indian English can be adopted.

6.5.6. *Ceylon or Sri Lanka English*

Although much has been written about Indian English, very little has been devoted to the subject of Sri Lanka (Ceylon) English. One article by Halverson (1966) bears mentioning. Because Halverson knew very little Sinhala and no Tamil, his study lacks the sophistication it would have had if he did. But he does raise a number of interesting questions. He notes, for example, that the phonology of C(eylon) E(nglish) has been much discussed, but that areas such as syntax, lexicon, and etymologies have been ignored. He cites the only serious study of Ceylon English done by H. A. Passé (1955) in which Passé does devote some space to the latter areas.

Halverson raises the age-old question of whether Ceylon English is a dialect of English or something else; his conclusion is that it is as much a dialect of English as American or Irish English, since it has native speakers, and because it innovates in style, lexicon, and syntax, something which, e.g. Danish or Swedish usages of English do not do (both in fact much closer to R. P. as a model). He notes that CE includes more than one dialect, many speakers controlling a sort of R. P. variety in some contexts, and reverting to a more relaxed CE type when speaking informally. Halverson points out that Passé ignores the probably strong influence of Tamil in CE, since the Tamils were earlier to learn English in Ceylon; Passé acts as if all the indigenous influences came from Sinhala.

Halverson points out some interesting stylistic peculiarities of CE, in addition to a number of idiosyncratic features, which he attributes to 'overgeneralization' (rather than to the structure of the indigenous languages, which he might have done if he knew more Sinhala or Tamil). The stylistic peculiarities he attributed to florid 19th century style, and also to non-standard (Cockney, Irish, Scots) usage. He shows in fact that simple modern English is often *not* understood, and his own use of the sentence 'If you need a place, I can get one' had to be translated into 'If it is a question of accommodation, suitable quarters can be arranged' by his CE-speaking intermediary.

Obviously the further study of CE needs to be undertaken by someone knowing the structure of Tamil and Sinhala to fill in some of the gaps in Halverson's work.

NOTES : CHAPTER 6

1. For further discussion of the role of bilingualism in ancient India see section 4.3.

2. For an excellent discussion of the problems inherent in analyzing problems of language census data, see Heinz Kloss (ed.), *Linguistic Composition of the Nations of the World,* vol. 1, "Introduction," Les Presses de l'Université Laval, Quebec (1974), pp. 3-42.

3. Thus, for example, Mangalorese and Madrasi are included in the lists of reported mother tongues in Gujarat in the 1961 census of India (Part II-C (ii)--Language Tables, p. 241).

4. For example, Islamic [presumably some form of Urdu], Gurmukhi [presumably Punjabi] and Kshatriya Gujarati are all included in the list of reported mother tongues in Gujarat (*Ibid.)*.

5. The major study to date of political manipulation of Indian language census returns is Brass, 1974. This work is useful in discussing the interpretations of language figures concerning Punjabi, Urdu and Maithili.

6. Mangesh V. Nadkarni, "Bilingualism and Syntactic Change in Konkani," *Language* 51:3 (1975), pp. 672-83.

7. Particularly Gumperz and Wilson, 1971.

8. Nadkarni, *op cit.*, p. 680.

9. Nadkarni, *op cit.*, pp. 680-1.

10. Nadkarni, *op cit.*, p. 681.

11. For a discussion of these earlier theories, see Haugen 1956:84 and Weinreich, 1953:119-21.

12. Cf. Das Gupta, 1969 and 1970.

13. This placing of primary emphasis on prestige has been shown to be overly simplistic in Emeneau, 1962a.

14. Gumperz, 1964 [reprinted 1971:207].

15. *Ibid.*

16. *Ibid,* pp. 207-8.

17. *Ibid.,* p. 210.

18. *Ibid.,* p. 217.

19. Gumperz and Wilson, 1971:256.

20. Gumperz and Wilson, 1971:251.

21. Pandit, 1972a:1.

22. Pandit, 1972a:4.

23. Pandit, 1972a:6.

24. Particularly the "Preface" by Dell Hymes (pp. 3-11) and the "Introduction" by David DeCamp (pp. 13-39).

25. Yet the work on "Swonals" (Speakers without native languages, cf. Tsou, 1975) shows that at least some individuals can be found who lose the ability to communicate in their mother tongue, yet never adequately acquire the second language of their environment. What is not clear is whether entire *communities* of swonals can be found, or whether 'swonalism' is a phenomenon restricted to individual or isolated (incomplete) bilinguals/swonals. Work by Bernstein (1964) on elaborated and restricted codes also indicates that some communities of speakers, particularly if illiterate or poorly educated, and of course socioeconomically depressed, may not have an elaborate linguistic code at their disposal, and may make do with a restricted one for many intents and purposes.

26. Cf. Schiffman, 1973:128.

27. Gumperz and Wilson, 1971:272.

28. Gumperz and Wilson, 1971:251.

29. Southworth, 1971:259.

30. For a general discussion of H and L forms in language, see Brown and Gilman, 1960.

31. Southworth, 1971:268.

32. Southworth, 1971:268.

33. Southworth, 1971:269.

34. Southworth, 1971:269.

35. Southworth, 1971:269.

36. Southworth, 1971:270.

37. Southworth, 1971:270-1.

38. (Schuchardt, 1889:508): "Die Sprachmischung plegt mit einer mehr oder minder starken Kulturmischung verbunden zu sein; mit der Blutmischung. . . sie wird. . . in keinem nachweisbaren Grade bestimmt. . . ."

39. Maria Theban (personal communication, 1975).

40. Ian Smith (personal communication, 1975).

41. Schuchardt, 1889:494.

42. Dalgado, 1900:xxiv.

43. Dalgado, 1900:3.

44. Dalgado, 1900:23-6.

45. In Dalgado's transcription, a non-italicized letter in an otherwise italicized transcription is apparently representative of retroflexion, while the opposite (one italicized letter surrounded by more non-italicized letters) is also representative of retroflexion.

46. Perhaps this means primarily Portuguese.

47. "Fazem parte de familia turânica" (Dalgado, 1917:40).

48. Thomson, 1961:109.

49. Thomson, 1961:109.

50. The Batticaloa dialect.

51. Chatterji, 1931:211.

52. Chatterji, 1931:212-13.

53. Chatterji, 1931:217.

54. Chatterji, 1931:219.

55. Apte, 1974:35.

56. Apte, 1974:39.

57. Sreedhar, 1974:71.

58. Sreedhar, 1974:71.

59. Sreedhar, 1974: 71 .

60. Since we have not been able to review many of these sources, please see Dharmadasa, 1974:80.

61. Dharmadasa, 1974:89.

62. Yule and Burnell, 1886:xv.

63. Spencer, 1966:62.

64. Southworth and Daswani, 1974:251.

65. Kachru, 1969:627.

66. Kachru, 1969:638.

67. Kachru, 1969:668.

68. Shanmugam Pillai, 1974:86.

69. Bansal, 1969:171.

Chapter 7

Ethnographic Semantics and the Ethnography of Speaking

7.0. Introduction and theoretical discussion

Part of the scholarly activity concerned with language and its interrelationship with society has been done partly or wholly within the discipline of anthropology. This work focusses on semantic systems such as those embodied in kinship systems, color terminologies, folk taxonomies of plants, animals, disease, or other systems in the environment of the culture bearers. Since its practitioners themselves do not agree whether to call this subdiscipline "formal analysis," "ethnoscience," "cognitive anthropology," "ethnographic semantics," "ethnotheory," and/or "ethnography of speaking," we also are hard pressed to label it correctly. Part of this difficulty may stem from the ambiguous position that language studies within anthropology are perceived to hold.

> The relation between language and culture seems a problem, it crops up whenever a thoughtful anthropologist tries to construct an integrated view of culture or behavior, yet discussion usually trails off irresolutely. We may set language and culture side by side, and try to assess similarities and differences; or we may try to see if something, a method or a model, that has worked for language will work for culture; or we may look to a future of point-for-point comparisons, once all partial cultural systems have been neatly analyzed; or we may redefine or subdivide the problem. We do not want to usher language out of culture; a suggestion to that effect some years ago was quickly suppressed. But having kept language within culture, many seem not very sure what to do about it. . . . (Hymes 1962:130)

Hymes points out that much linguistic analysis has not fit with an analysis of culture, because the techniques devised for linguistic analysis have been too abstract for analyzing what is pertinent to both; similarly many approaches to the understanding of culture have ignored a large body of activity which falls between the two fields. Furthermore, "speech as such has been assumed to be without system; its functions have been assumed to be universally the same; the object of linguistic description has been assumed to be more or less homogeneous; and there has been an implicit equation of one language = one culture." (Hymes 1962:131)

Thus there exists a large amount of linguistic activity which has been ignored or at least relegated to peripheral status by both linguistics and anthropology, which Hymes would like to see subsumed under a subdiscipline he calls the "ethnography of speaking."[1]

"Speaking, like language, is patterned, functions as a system, is describable by rules."[2] It is the purpose of Hymes' article to call attention to those areas of language activity which in fact exhibit structure, albeit not homogeneity, and claim them for anthropology and its new subfield, the ethnography of speaking. Thus he proposes for us "to take as a working framework: 1. the speech of a group constitutes a system; 2. speech and language vary cross-culturally in function; 3. the speech activity of a community is the primary object of attention."[3]

It is our goal in this chapter to focus attention on this area of what used to be considered exclusively anthropological concerns and to examine what has been done in relation to South Asia. Since there has been a great deal of attention paid to kinship, we devote a fair amount of space to such studies; studies which Hymes would call now "ethnography of speaking" are given another block of space at the end of this chapter (7.2); the somewhat unfortunate break between the two kinds of concerns is probably related to the historical development of such concerns--the study of semantic systems or domains other than kinship seems to have developed out of the kinship model and as far as our material is concerned, postdates much of the kinship work. Yet new approaches to kinship are constantly being made, so that earlier studies reflect earlier concerns, and later studies fit more closely with non-kinship studies. The terminological difficulties are still in evidence, since we have discussed some articles, which others might consider to be clearly ethnographic semantics, under other rubrics in this volume.

7.1. *Ethnographic semantics*

7.1.1. *Kinship*

7.1.1.0. *Introduction*

The place of kinship with relation to language and culture in South Asia is discussed and summarized by Southworth and Daswani (1974:190-5) in their text *Foundations of Linguistics:*

> What is the nature of the relationship between the system of kinship terms and the culture, or (since the kinship system is part of the culture) how does it correlate with other kinds of cultural behavior? . . . It is often assumed that, if a distinction is made consistently by people in a society, that distinction must be somehow relevant to them. . . . Generally speaking, it is reasonable to assume that where distinctions such as that between FEB, FYB, MB, etc. . . . exist, that these different categories corresponded to different role expectations. An example of this may be found in the South Indian kinship systems, in which one's parallel cousins. . .are generally called by the same terms as one's own brothers and sisters, but cross cousins. . . are called by distinct terms. . . . This distinction would appear to correlate with the fact that the preferred marriage partner in this area for a male is the younger cross cousin [FZS or MBD], whereas marriage with parallel cousins is not permitted. In North India no marriage with a cousin is permitted. Thus in both cases, the terms for *brother* and *sister* designate individuals whom one cannot marry.[4]

Southworth and Daswani thus show that while the English system differs from both the Hindi-Urdu system and the Tamil system, the latter two systems also differ. They point out that terminology and system are not always congruent--that the presence or absence of terms does not necessarily signify difference in social behavior. But where differences in linguistic terminological distinctions do correlate with social behavior, they claim that a difference in *role expectations* is to be found. Much is made of this in studies we will examine later, some of which we will in fact not examine in great detail. We will try to concentrate in this section on studies which are firmly grounded in linguistic analysis, but which also go beyond terminological distinctions to posit semantic systems which are firmly correlated with both linguistic distinction and social behavior. Much of the literature on kinship breezes past the linguistic distinctions and prefers to construct philosophical arguments based on ideal systems, seeking refuge in the linguistic terminology every time some philosophical construct is found to be weak. Others concentrate on the linguistic terminology to the exclusion of any correlation with social behavior. Few studies in fact wed the two successfully, which seems to underlie the justification for Hymes' lament quoted in the first paragraph of this chapter. We will, however, mention studies embodying both these weaknesses, in order that the reader might have some idea of what work has been attempted on the subject, and might follow some of the arguments presented in various studies.

In surveying the field of kinship studies related to South Asia, it is necessary to consider a number of studies of a somewhat theoretical nature which do not necessarily specifically mention South Asia, but which are responsible for the formation of some of the basic concepts of the *formal approach,* as opposed to more traditional kinds of kinship studies of the *structural* type, which are aimed at finding structure in society and pay only lip service, if at all, to linguistic structure.

7.1.1.1. One of the basic studies of the former type (despite the word "structural" in the title) is Lounsbury, 1964, "The Structural Analysis of Kinship Semantics." This article is important because of his rigorous definition of the terminology involved. He points out that the lexical sets involved in a kinship terminology constitute a paradigm, and can be analyzed like other paradigmatic sets in a language. He goes on to state that linguists consider kinship vocabularies and their meanings as something special in lexicology, permitting the kind of rigorous reference they do, but find them unrepresentative of linguistic/semantic, or lexicological problems in general. Lounsbury feels that there is something a bit special about the structure of kinship systems, precisely their paradigmatic nature. ". . .In the perfect paradigm, the features of any dimension combine with all of those of any other dimension. In the perfect taxonomy on the other hand, they never do; they combine with only one feature from any other dimension."[5] Lounsbury also points out an important distinction about the Dravidian systems, which are not generally founded on clan or moiety reckoning, but on "a mode of reckoning of bifurcation that, unlike the Iroquois takes account of the sexes of all intervening links."[6] This feature has interested and plagued many scholars working on the Dravidian systems, and as we shall see, remains unsatisfactorily treated by most of them.

7.1.1.2. Wallace and Atkins, Meaning of Kinship Terms

Another important article in the development of formal analysis is that of Wallace and Atkins (1960). In their conclusion, they state that "semantic analysis in anthropology concerns, primarily, neither personality and culture, nor linguistics, nor culture and social structure per se, but cognitive processes in culturally organized behavior."[7] This explicit definition of the domain of semantic analysis being that of cognitive processes parallels closely the development of the latter-day emphasis on linguistic competence as a mental process distinct from performance. It also recalls the time-wearied concern in linguistics with the psychological reality of the phoneme. They emphasize that what formal analysis (semantic analysis) is concerned with is the reality of semantic structures as the culture-bearer perceives them; "The psychological reality of an individual is the world as he perceives and knows it, in his own terms; it is his world of meanings."[8] What Wallace and Atkins want to avoid is the structural reality applied to a society by the ethnographer; the world of meanings applied to a society which is real to the ethnographer is not the same as the meanings which are real to the culture-bearer. Throughout this section we shall see that this polarity between the two realities is obvious in much of the work on kinship.

7.1.1.3. Frake (1962) then carried the concern for finding conceptual systems to areas other than kinship: ". . .all peoples are vitally concerned with kinds of phenomena other than genealogical relations; consequently there is no reason why the study of a people's concepts of these other phenomena should not offer theoretical interest comparable to that of kinship studies.[9] For an example he gives Brazilian Indians who classify birds and parrots in a certain way:

> Culturally significant cognitive features must be communicable between persons in one of the standard symbolic systems of the culture. . . A major share of these features will undoubtedly be codeable in a society's most flexible and productive communication device, its language. Evidence also seems to indicate that those cognitive features requiring most frequent communication will tend to have standard and relatively short linguistic labels.[10]

Frake apparently intended that the guidelines he was trying to establish would constitute a set of field procedures which would "provide the ethnographer with public, non-intuitive procedures for ordering his presentations of observed and elicited events according to the principles of classification of the people he is studying."[11] "The principles by which people in a culture construe their world reveal how they segregate the pertinent from the insignificant, how they code and retrieve information, how they anticipate events, ...how they define alternative courses of action and make decisions among them."[12] So giving "a central place to cognitive processes of the actors involved will contribute reliable cultural data to problems of the relations between language, cognition and behavior,"[13] and this will give us descriptions which will "succinctly state what one must know in order to generate culturally acceptable acts and utterances appropriate to a given socio-ecological context . . . ."[14] Here again the emphasis is on studying formal systems in a culture through the paradigms of the language that reflect how the members of the culture think about and structure their world.

7.1.1.4. Cognitive Anthropology

The formal analytic approach is summarized by Tyler (1969), who uses the term "cognitive anthropology" to cover the same ground as ethnographic semantics, formal analysis, ethnoscience, ethnosemantics, etc.

> . . . cognitive anthropology constitutes a new theoretical orientation. It focuses on *discovering* how different peoples organize and use their cultures. This is not so much a search for some generalized unit of behavior analysis as it is an attempt to understand the *organizing principles underlying* behavior. It is assumed that each people has a unique system for perceiving and organizing material phenomena--things, events, behavior, and emotions (Goodenough 1957). The object of study is not these material phenomena themselves, but the way they are organized in the minds of men. Cultures then are not material phenomena; they are cognitive organizations of material phenomena.

"In essence, cognitive anthropology seeks to answer two questions: what material phenomena are significant for the people of some culture; and how do they organize these phenomena?"[16] Tyler also comes out squarely in favor of studying *variation* in semantic domains, rather than assuming that variation (cf. free variation in linguistics) is insignificant. "A consequence of this interest in variation is the idea that cultures are not unitary phenomena, that is, they cannot be described by only one set of organizing principles."[17] "Variations are not mere deviations from some assumed basic organization; with their rules of occurrence *they are the organization*."[18]

Tyler also makes it clear that it is through *language* that the anthropologist, cognitive or otherwise, must seek the native categories: "thus far, it has been assumed that the easiest entry to such processes is through language, and most of the recent studies have sought to discover codes that are mapped in language. . . .how other peoples 'name' the 'things' in their environment and how these names are organized into larger groupings. . . . Naming is seen as one of the chief methods for imposing order on perception."[19] "It is through naming and classification that the whole rich world of infinite variability shrinks to manipulatable size and becomes bearable."[20]

After stating the different kinds of organization of named things (semantic domains) that languages exhibit, and how these organizing principles differ (taxonomies, paradigms, trees) and the necessity of having techniques to discover these domains, Tyler points out that there is more to the semantic system than what is communicated and how it is communicated. "Other semantic features deriving from the context of communication are equally important. Context includes the manner of communication (for example, verbal and written), the social setting, and the linguistic repertoires of speaker and hearer. Contextual semantic features and their mutual interdependence are as much a part of the cognitive system as taxonomies and semantic domains."[21] Advocating that anthropologists must go beyond the lexeme because of the properties of categories being either perceptual or conceptual, Tyler states that ". . . In connected discourse speakers and authors deliberately manipulate semantic features in order to convey nuances of meaning often quite opposed to the overt content of individual lexemes."[22] Tyler even goes so far as to claim that the preoccupation with lexemes may obscure historical reconstructions, because lexical items may change even though the system may remain the same. Abandoning the concern with the lexical items allows us to reconstruct semantic systems, for example the Proto-Dravidian kinship system, by concentrating on the categories and ignoring the inconsistencies in the lexical representation of these categories.

Having examined some of the historical background of ethnographic semantics, so to speak, through the works of Lounsbury, Goodenough, Wallace and Atkins, Frake, Hymes and others to the explicit goals stated by Tyler, we will now examine the work on semantic domains (primarily kinship studies but also some work on other domains) that has been done on South Asia. Many of these studied antedated the development of ethnographic semantics; many of them are concurrent, but may be ignorant of, indifferent to, or even hostile in its concerns. Because of the many different approaches and the great confusion existing between and within the directions of various scholars (witness for example the debate between Tambiah 1965 and Yalman) we will attempt to hew closely to a policy which holds that language tells us something about the culture, and only those studies which keep both language and culture clearly in view will be considered by us. There is unfortunately much in the literature, especially in kinship studies, which ignores this connection, and we judge these studies harshly, especially when they claim to be concerned with explaining the culture through language while ignoring the categories so obviously present in the linguistic system.

7.1.2. *Structural Approaches to the Study of Kinship in South Asia*

Typical of the studies which have difficulty in keeping the connection between language and culture straight are the so-called 'structural' approaches, some of which are heavily French Structural à la Lévi-Strauss. Although their authors profess to be concerned with the 'categories' or 'terminologies' of the kinship systems, they seem to make intuitive leaps direct from the terminologies (which they often in the same paragraph refer to as categories and/or rules as well) to the modes of behavior of the culture bearers without pausing to examine how the kin-terms differ from the categories that they represent, and how rules of behavior are different from both of these.

7.1.2.1. Thus Yalman (1967) can say ". . .the Sinhalese terminology does play an important part in its kinship system [sic!], and it raises significant problems about the regulation of sexual relations within the family and the nature of the concepts of incest and exogamy."[23] And further "I prefer to treat kinship terms as 'categories' whereby the Sinhalese organize their kinship universe; these categories are definitely associated with certain rules of behavior."[24] Yalman presents us further with a list of kin terms, and then proceeds to refer to this list as the categories themselves: "a cursory examination of the categories will indicate. . ."[25]

Yalman seems to accept the notion, which he attributes to the Sinhalese, that *the word is the thing;* that is, the terminologies are the categories. In other words, Yalman takes the approach of getting at the conceptual categories of the culture bearers by accepting *their own* analysis of what words mean, a trap that better linguistic training might have saved Yalman from. Yalman also summarizes the whole system in one 'rule' which encompasses the whole of Sinhalese kinship:

> What then, is served by these elaborate categories, and why are the Sinhalese villagers so strong in their support of them?. . . The Sinhalese, like all other peoples in the world, must somehow restrict and channel sex relations in kin groups. They do not use 'exogamy' or any other alternative sets of specific prohibitions. They use a single positive rule: that the only persons who may legitimately have sex and may marry are those standing in the prescribed categories of *massina-nana*.[26]

"The rule of massina-nana unions, . . . must be seen as a corollary to caste endogamy intended to restrict and specify the legitimate sex mates within the castes."[27] This attempt to summarize all Sinhalese kinship behavior into one 'rule' may be elegant in its characterization of Sinhalese marriage behavior, but it oversimplifies the relationship between the categories of the system and its terminology.

Yalman also seems to completely misunderstand the approach of the ethnographic semanticists; he attacks Tyler, for instance, for being a behaviorist, which is probably equivalent to calling Chomsky a Skinnerian. For example: ". . .it will take much persuasion to wean most gifted ethnographers from their behaviorist modes of thought and expressions"[28] and in a footnote, he specifically attacks Tyler:

> S. A. Tyler, for instance, appears to deny the relevance of the Dravidian forms so meticulously outlined by Dumont. He seems to think that because of the great variation in many diverse contexts of the actual usage in address of the kinship terms, there is no unitary structure to the terminology as such. All that can be done, it appears, is a meticulous analysis of all these contexts which determine the expressions. This is an extreme behaviorist position. . ."[29]

This seems to be a misreading of Tyler, since Tyler would presumably not overlook "the semantic power of the terms and what this means for the structure of the cognitive pattern, sexual prohibitions, and positive marriage rules which are connected precisely with the 'form' of the terminology itself" (Yalman 1969:625). What Yalman seems to misunderstand is that Tyler feels that the structure is in the *system*, not in the terminology, and even when the latter is confusing, the system *underlying* it is not. Since Yalman seems to think that the structure *is* the terminology, everyone who denies this is dismissed with the most convenient label available, i.e. 'behaviorist.'

7.1.2.2. Since discussions of South Asian kinship systems in general and Dravidian systems in particular always bring up the name of Dumont, we need to examine what he has contributed to the study of kinship. The title of his article "The Dravidian Kinship Terminology as an Expression of Marriage" (Dumont 1953) reveals already a possible confusion such as that seen in Yalman, between the terminology and the system. Dumont fortunately does realize that there is at least one step between the terminology and marriage, i.e. he recognizes categories of classification according to generation, distinction of sex, of two kinds of relatives inside certain generations (elder vs. younger siblings), a distinction of age, etc. But interestingly, Dumont never explicitly gives us the terminology; he has worked out the system from the terminology, apparently, but we are never presented with how he arrived at this system. The data are withheld, and we are forced to work with his system only. It is therefore impossible to either verify or falsify his constructs; we either accept them or risk being branded a 'behaviorist' by his disciples. Since it is impossible for us to tell how Dumont has arrived at his theoretical construct, it is difficult to criticize it here, and we will refrain from doing so. This is not to say that his theory lacks validity; but since it is our stated objective to keep always in mind the connection between language and culture, his study lies outside the scope of this chapter since it presents us with no language material.

7.1.2.3. Another study which attacks Yalman, but misses a number of positive points is that of Tambiah, "Kinship Fact and Fiction in Relation to the Kandyan Sinhalese" (1965). Tambiah's criticism of Yalman is that Yalman adheres too closely (!) to the terminological system and ignores inconsistencies in it. That is, Yalman (according to Tambiah) assumes that the terminology and the practice it implies are the same, i.e. 'congruent.' When this congruence is clearly not present, i.e. when cross-cousin marriages are the exception rather than the rule, Yalman supposedly retreats into the terminology. Tambiah's contribution is to show that the 'facts' are quite different from the 'fictions' of the Kandyan system of kinship--the system ideally predicts one kind of behavior, but people deviate from that ideal system too often for us to be able to accept that system. "The kinship ideology of the Kandyan Sinhalese contains two fictions. The first fiction is that of *gedera* implying a group, agnatically recruited by virtue of *dīga* marriage and tied to a locality by virtue of ancestral property. . . .The second fiction is that of endogamy in both its aspects. . . ."[30] According to Tambiah, ". . .extra-kinship variables, primarily economic in nature"[31] are responsible for the deviation from the ideal system. This is all very well and good, and Tambiah's criticisms are perhaps valid, if Yalman's constructs are a model of behavior. But Yalman's constructs, as far as we can see, are a model of 'competence,' rather than 'performance.' We have already levelled criticism at them for confusing form and function and meaning; but Tambiah seems to misunderstand what Yalman's model is all about. We would expect Yalman to level the 'behaviorist' label at Tambiah, rather than at Tyler. Due to the various confusions and misunderstandings rampant here, we will not attempt to resolve the conflict between Tambiah and Yalman, but say only that a finer attention to the facts, i.e. the details presented by the language itself, might obviate some of these problems.

7.1.3. *Data-Oriented Systems of a Non-Formal Type*

Another kind of study, which is more data-oriented, but does not attempt to extrapolate semantic systems from the data, is found in the works of Emeneau (1941, 1938) and Bhattácharya (1970). We include the last primarily because it is the only study available on Munda kinship.

7.1.3.1. Emeneau, in the Coorg study, points out the fallacy of expecting to be able to predict social behavior from the kinship system.

> It seems unwise. . .to base a great deal on the evidence given by kinship terminologies in the Dravidian-speaking areas of India. They are practically identical in all the languages of the area, and the same terminologies are used by such ethnologically diverse communities as the mother-sibs of the

> Malabar coast, the father-sibs of most of the rest of the area in question, the Coorgs and the Todas who practise symmetrical cross-cousin marriages and communities which practise an asymmetrical form of cross-cousin marriage. Until we have exact ethnological and linguistic accounts of many more communities of the area than we have now, any correlations found between institutions and terminology must rest under suspicion of being in part accident and not due to a coordinated development of the two.[32]

Thus while giving a careful analysis of the kinship terminology and the system it represents, Emeneau shies away from predicting behavior on the basis of it. This is the dilemma which inevitably proceeds from the thesis that performance ought to be predictable from competence. Emeneau's work was done at a time when this distinction was not emphasized in American linguistics.

Emeneau's other article of this type is that on Toda kinship terms (1941) which is comprehensive but ethnographic rather than attempting any semantic analysis.

7.1.3.2. The article by Bhattacharya (1970) is very descriptive and contains probably the fullest data on Munda available. In its conclusion he attempts a brief analysis of some of the categories involved: "in the kinship system of those tribes, status is more important than sex or any other consideration. . . It will. . . be noticed from the kinship forms and their meanings. . .that the Munda kinship terms as a whole are highly classificatory in nature extending over different generations, sexes and lineages."[33]

Bhattacharya is also interested in the fact that typologically certain classifications show up over a wider area than typified by one language group, i.e. that there is a 'kinship area' in Central India which crosses language boundaries; but he also warns against making predictions based on the system, when it may be out of date and no longer reflecting actual practice.

7.1.4. Formal Analysis Studies Proper

In this section we finally can isolate a number of studies which meet the ideal notion that we have set up for this section--studies which skillfully wed the study of language with the study of culture and do not stray from this ideal.

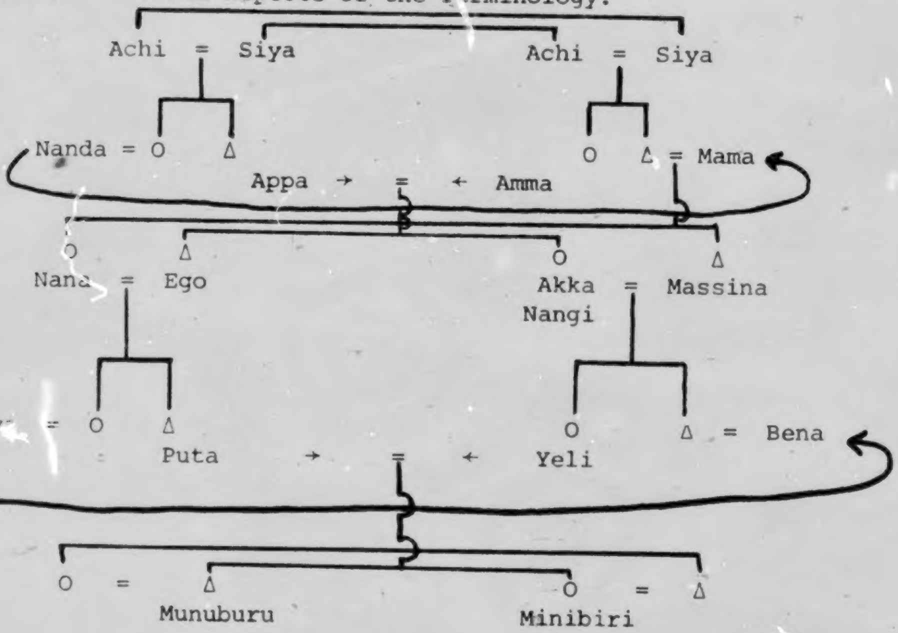
To do justice to Yalman, we include his article on "The Structure of the Sinhalese Kindred: A Re-Examination of the Dravidian Terminology"

(Yalman 1962) here, since he succeeds more admirably in keeping the terminology and the categories and behavior connected with it separate. He does not completely succeed, but it is easier to see what he might have said, had he succeeded, than in other works.

7.1.4.1. In what may be a thinly veiled attack on Tambiah, Yalman states that "The Sinhalese rules regarding sex and marriage are entirely dependent upon the terminology of kinship. . . . They are impossible to comprehend without taking the terminology into account; . . . ."[34] While he also confuses terminology with categories and rules ("The hypothesis I propose is that these abstract rules are not [etc.]. . . but that they are systematic categories which form the internal structure of bilateral kindreds of the Sinhalese type."[35] Yalman manages to construct a diagram for us however, which quite elegantly illustrates the "Formal Aspects of the Terminology"[36] which we reproduce below.

His main point here seems to be to show that it is the categories which specify the "Correct marriage partners for all persons in a kin group"[37] rather than exogamy, which has been given as the explanation by many others. If Yalman were clearer in his distinction of terminology, category, and rule, it might be possible to judge whether he has proven his point.

Figure 61: "Formal Aspects of the Terminology."[38]



7.1.4.2. In Tyler's article "Koya Language Morphology and Patterns of Kinship Behavior" (1965) an attempt is made to relate the pragmatics of Koya kinship term usage to the social employment of the terms, by examining the relationship between language structure and pragmatics. "The particular relationship under discussion is that obtaining between *morphology* and *normative* rules of respect-intimacy behavior within the domain of kinship behavior and linguistic reference to kinsmen." (Tyler 1965:1428) The formula for the Koya kinship term consists of:

Possessive	±	Qualifier	+	Stem	±	Derivative Pronominal Suffix	±	Plural Suffix

The morphemes which are lexically realized in these constituent slots are:

maa- naa-	±	chinna- beri- meena	+	Stem	±	-aal -Du -Di	±	-ooru -ku -n -sk

Then Tyler shows that there are four patterned behavior types: respect, reserve, informal, and intimate, and that they are correlated with features of linguistic usage. However, there are some variations in usage, and Tyler is not able "to establish a situational context for the alternation"[39], although he feels that "functionally, these variations seem to be related to contradictions, inconsistencies, or 'play' in the roles of these relatives."[40] After indicating that languages which mark respect/intimacy in the morphology rather than in the lexicon are Dakota-Iroquois systems and/or "societies with some tendency to lineage organization,[41] Tyler proposes that "kin types whose role systems are characterized by a high degree of inconsistency will have a higher frequency of alternate lexemes or morphemes indicating degrees of respect-intimacy than those kin types whose role systems are more consistent."[42] Thus treating variants as synonyms would not capture the generalization Tyler is trying to posit between variation and role system inconsistency. This study more than any other we have examined explicitly tries to show a correlation between language structure (morphology) and respect-intimacy behavior. Whether Tyler's generalizations can be shown to hold true for other languages in South Asia remains to be seen.

7.1.4.3. In another article (Tyler 1966) Tyler addresses the problem of the *form* of kin terms varying according to specific social context. Tyler again attacks the notion proposed by Leach that variation can be discounted as a kind of synonymy. Tyler feels that concern with the lexemes alone ignores the whole contexts in which terms of reference are used: "Elsewhere I have argued that a part of the variation occurring in kinship terminologies can be explained by the contradictions entailed in the role systems of the class of kinsmen denoted by a given term. . . . In this paper I want to explore further implications of role specificity for variations in terminology. . . .to relate terminological variation to the contexts in which terms of reference are used."[43] Aside from phonological variations, there is also great lexical variation in the Koya's use of kin terms; they can use Telugu terms or Koya terms. "The problem would be simple if a bilingual Koya used items from list one [Telugu] when speaking Telugu and items from list two [Koya] when speaking Koya. Isolating the terms in this fashion, however, is misleading, for all forms occur in the linguistic repertoire of a single speaker."[44]

Tyler shows that the semantic distance between the two codes is relatively minor. "A componential analysis reveals the structure of the two codes. . . .The components are (1) sex; (2) generation; (3) cross vs. parallel; (4) relative age."[45] In most respects the two systems are coordinate, hence code-switching or 'transforming' from one system to the other is fairly simple, and the difference can be characterized by a few simple rules that are part of a speaker's competence in switching, e.g.

$$1.\quad \left\{\begin{matrix} a+m+m \\ a+f+m \end{matrix}\right. \rightarrow \quad a+a+m$$[46]

These can be simplified further in various ways; Tyler also specifies the environments where some other variations occur, e.g. whether the Koya is Christian or not, speaking to a Telugu Christian or not, etc. "In general, at the level of semantic structure these. . .transform rules plus the accompanying distributional statement enable one to predict the appropriate denotata of a term."[47]

But there is a second type of lexical variation relating more to situational *variants within a kinship code* than to switching codes. For instance there are formal/informal or respectful/nonrespectful usages; there are intimate or 'sweet' terms used only in the home between consanguineal kin; there are also different possessive pronouns and derivative suffixes which he tabulates both for Koya and Telugu (the previous article by Tyler is concerned with these phenomena); in Koya only there are further conditions; 1. sex of speaker; 2. audience composition, involving two factors; presence/absence of the referent; and presence/absence of speaker's elder consanguineal relatives and/or presence/absence of nonkin. The major events where this is important are certain weddings and festivals. Since Christians no longer use these, the difference between k (Koya Christians) and K (non-Christian Koyas) is noted, which Tyler gives a rule for. Furthermore, between grandparents and grandchildren there is an intimate relationship, so younger children use the pronoun /naa-/ for these relatives (immature speaker inside a house). Also status differences between speakers are reflected in the use of mii/nii possessive pronouns for 'your.' With derivative suffixes, the variable is *audience composition:* the presence/absence of elder affines.

In conclusion, Tyler states that however ". . .sketchy and incomplete. . . this analysis is, it should at least indicate that the appropriate use of Koya kin terms cannot be predicted solely on the basis of a formal analysis predicated on the assumption of genealogical reckoning. There are many contextual factors to be taken into consideration. . . . most difficult of all [of these is] something that might be called the speaker's intention."[48] Thus Tyler in choosing to grapple with variation has broached the same kind of problem handled by Labov and others in variability studies in English and other western languages. (Labov 1972a) It is clear that this is an area where the greatest amount of work needs to be done in South Asian languages, and the contribution towards a beginning of variability studies in ethnographic semantics has been competently made by Tyler in this article.

7.1.4.4. A final study in this section of formal analysis papers is that of Leaf (1971). He attempts to bridge the gap between structural and formal analyses by using the terminology of the Punjabi kinship system as the labels of the model of the system he attempts to inductively generate. This seems to be similar to attempts in the past to represent phonemes by the letter most commonly associated phonetically with that phoneme, e.g. the inventory of English phonemes includes members "called" /a/, /i/, /u/, etc. To achieve this Leaf proposes an elicitation procedure which involves members of the culture checking and agreeing on names for spaces in the system at each point. This is based on the understanding that a "semantic system is a system of related definitions and that terminologies embody such systems. To do so permits terminologies to be seen as complete linguistic systems, but avoids the implications of kin types."[49] The terminological labels are thus not kin types, nor are they meanings, but 'links' between people in genealogies. Leaf also shows how the linguistic pattern supports the semantics of the system through the morphological devices available to distinguish different kin. In sum, the object of this study is to determine cognitively distinct--for the culture--semantic spaces and what "phonological words" fit in the spaces. A "word" which is associated with a "space" may by extension, be associated with some features of that space; thus, e.g. the features of age, male gender, and emotional closeness may be identifed by referring to a non-biologically or socially related individual as "grandfather."

To achieve this rather unambiguous system of labels that Leaf sets up, he unfortunately commits the mistake, so common in studies of these kinds, of lumping all "synonyms" together as one "word": all forms with the same definition are the same "word", and the phonological shape of a particular "word" has no intrinsic or inherent relation to its semantic aspects. ". . . A procedure appropriate to this sense . . . [represents] a single unambiguous definition."[50] As we have seen from Tyler's study (1966) this is a very dangerous technique, since many subtleties and in fact real "meanings" of the kinship terminology are lost by overuse of the concept of synonymy. Leaf's deceptively simple and elegant system, which is also supposedly that of the culture bearers rather than that of the ethnographer, is thus achieved at the expense of ignoring cognitively and socially real differences.

7.1.5. Applied and Comparative Studies

In this section we examine studies of kinship that are what we call comparative and applied--the latter in particular are straightforward applications of one kind of kinship theory or another to some culture, occasionally with some theoretical implications, but in general theory is not the main thrust of the article. In this they differ from the previous section where theoretical considerations overrode the practical implications of the study. We also examine several studies which are in fact comparative--examining a number of systems or subsystems in an area, and drawing comparative conclusions.

7.1.5.1. Vatuk's "Structural Analysis of the Hindi Kinship Terminology" (1969b) is a solid analysis of the Hindi system, a complete classificatory and descriptive study that in several points quibbles with Dumont, but in general leaves few stones unturned.

7.1.5.2. Tyler's "Koya Kinship Terminology" The Relation Between Syntactic and Semantic Analysis" (1968) is a formal analysis of Koya kin terms with a brief theoretical argument concerning multiple analyses. Parts of this analysis have appeared in other studies by Tyler; of interest here is an attempt to see whether syntax and semantics are correlated. He concludes that ". . .the morphological analysis points up a tendency for generational components to be rather weakly realized in most Dravidian kinship systems." (Tyler 1968:358)

7.1.5.3. A parsimonious study by Khokle of kinship terminology among the Deshastha Brahmans of Maharashtra[51] very rigorously outlines the terminology, the morphological analysis isolating various phonological material which correlates with generation, descent, etc., followed by a componential analysis isolating conceptual variables. Aside from the presentation of the formal analysis, little discussion accompanies this otherwise thorough study.

7.1.5.4. Among comparative studies is M. Shanmugam Pillai's "Caste Isoglosses in Kinship Terms", (1965a) in which he has elicited kinship terms from a number of different Tamil castes (and Muslims) and compares shared vs. unshared terms. He proposes that groups that share the most number of terms are the least isolated, while groups such as the Muslims are the most isolated because they have the highest percentage of different terms. This is essentially a kind of dialectology of kinship terms, which the title of the article clearly shows. It remains to be seen whether his thesis is confirmable.

7.1.5.6. Another article with interesting comparative data is that of Gough (1956) which is primarily an ethnographic description of the traditional Brahmin family structure. But she also gives a good deal of data about non-Brahmin terminology, and contrasts the two systems.[52] She draws the conclusion that the social relationships in Brahmin families are different from those in non-Brahmin families because of differences in importance of land, ritual sanctity, goals, inheritance, position of women, attitude toward sexuality, etc.

7.1.5.7. A final article in this section is that of Kumaraswami Raja (1972) which compares the Tirunelvēli Tamil dialect's 'personal' kin terms with the Kshatriya Rajus' Telugu dialect of Rajapalayam. Since there is evidence that the fused construction involving kinship terms (OTa. has /em -pi/ 'my younger brother', /num-pi/ 'your younger brother' etc.) is a Proto-Dravidian feature,[53] Raja raises the question of why this feature is preserved only in Tirunelvēli Tamil and Kshatriya Raju Telugu, but in neither standard Tamil nor standard Telugu. Raja points out that the two dialects are not distant in space, and wonders whether this is evidence for a kind of shared "semantic" system. It is not clear to us that this involves shared semantics, although some semantic features are involved. Perhaps what is happening here is that there is a phonological rule for the fusing of pronouns to the kin term, which correlates with terms for older relatives only (a semantic feature); what is shared is a rule, perhaps, rather than a semantic system.

7.1.6. Fictive Kinship

The final section in the studies on kinship is devoted to a number of interesting studies on fictive kinship--the use of kinship terms for addressing or referring to individuals who are not real kin.

7.1.6.1. The first of these, by Stanley Freed (1963) lays the groundwork for the second, by Vatuk. His conclusion which summarizes his findings aptly, needs only to be prefaced with a statement of the facts of the situation, namely, that in the village of Shanti Nagar that he studied, villagers use kinship terms to practically all of the members of the village, this being defined as people settled in that village longer than anyone's memory, or related in a true way to such people. Villagers who are recent immigrants are not accorded the use of the kinship terms.

"This paper, in presenting an analysis of the use of fictive kinship terminology in Shanti Nagar, attempts to demonstrate that fictive kinship terms are determined primarily by (1) the fictive genealogical system of the village and (2) considerations of relative prestige due to the caste hierarchy."[54] This means that villagers assign others to a hierarchy of generations based on fictive genealogies, i.e. "your father and my father are like brothers, so we are like cousins."

> Comparison of elicited fictive kinship terms with the fictive genealogical system shows 81 percent of agreement with the terms to be expected from fictive genealogical connections. Examination of the terms not in agreement shows that some respondents systematically adjust their positions up or down a generation with regard to particular castes or lineages. All such adjustments appear to be attempts on the part of individuals to bring their positions into agreement with what they feel their status to be with reference to other persons. In most cases of terminological adjustments, respondents of higher castes raise themselves a generation with respect to lower castes . . . and respondents of lower caste lower themselves a generation. When these adjustments are taken into account 89 percent of the fictive kinship terms can be satisfactorily explained. The remaining 11% of the terms appear to be respondents' errors. The evidence presented strongly supports the hypothesis that fictive kinship terminology depends upon the fictive genealogical system and considerations of the relative prestige of castes.[55]

7.1.6.2. Vatuk's article (1969b) departs from Freed's earlier approach. She studied the use of fictive kinship terminology in an urban *mohallā* where most people were in-migrants and had no long-standing kinship ties. Yet most residents of the *mohallā* used some kind of kinship terms fictively for address and reference, and the strategies they devise to decide which terms to use are extremely interesting. "An analysis of fictive kinship usages in the urban *mohallā* shows that there exists here no internally consistent fictive genealogical system comparable to that described by Freed and recognized by urban residents to have existed in their home village. . . .such a system exists in the older *mohallās* in the central part of the city. . . .But in the new *mohallās* described here such a concept has not evolved." But courtesy alone will not explain what goes on. The choices in the Hindi terminology are wide and the choice reveals "much about the traditional social structure and the direction of change under urbanization."[56]

Vatuk begins her analysis by describing the kinship terminology and how it is used by real kin. She notes that given names cannot be used in the *mohallā* because use of given names implies the speaker is superior to the addressee. "Ma" and "Mas" are not used and "shrimati" cannot be used without a given name, and is only used anyway with highcaste outsiders.

Vatuk points out that people in the *mohallā* always begin by asking people how they are related to each other, or to themselves. When this is determined, they can then assign the proper kin term, then people are not related, but friendly, it is considered rude to say there is no relationship; instead they will say something like "We say 'sister' to one another" or "I consider her my sister."[57] People thus can be classified in a range from real kin to possible kin to fictive kin. "The choice among them is one for the individual to make, but it is not made randomly."[58]

The area where fictive kinship in the *mohallā* has its most extensive application is in *address*. In the traditional terminology there is a distinction between ***sasurāl*** and ***pīhar*** kinship. ***Pīhar*** is the mother's village, and the ***pīhar*** kin are her cognates. The ***sasurāl*** is the husband's village, where the wife takes up her residence. The ***sasurāl*** kin are her affines. Thus the residents of the *mohallā* can be addressed fictively with either the ***sasurāl*** or the ***pīhar*** kinship terms, which appears to be a very confused situation. "One notes inconsistency in the use of terms. . ."[59] "A man may refer to a male friend as "elder brother," but to the latter's wife as 'elder sister' rather than "elder brother's wife." From these and other examples one might conclude that kin terms are chosen more or less at random,"[60] and that sex and relative age are the only bases on which to develop fictive kinship.

For most residents of the *mohallā* it is neither a ***sasurāl*** nor a ***pīhar*** kind of place, although it is rare that an individual does not have some kind of kinship ties (probably very distant) in the *mohallā*. Vatuk claims that there are two important variables which determine the neighborhood pattern of fictive kinship--an egocentric approach, actually a composite of overlapping egocentric fictive kindreds--but inclusive, interlocking network of fictive links. The critical variables are: (1) childhood residence of Ego or his spouse in present or adjoining *mohallā*. (2) existence of real or 'village' kinship ties between Ego/spouse and other *mohallā* residents. That is, there are a number of ordered choices which a resident uses to establish fictive kinship--the first question is, is the *mohallā* home for the man, and thus sasural for the woman; secondly is the *mohallā* pīhar for the woman (which would make the *mohallā* uxorilocal for that family, which there are actually no true types of in these *mohallās*); thirdly, having established neither (1) nor (2), the residents are free, if they are neolocally resident in the *mohallā*, (which is the most common case), to establish fictive kinship on the following basis: "In this case they use for any real or 'village kin' the appropriate terminology and for other neighbors with whom they become acquainted choose kin terminology which sets up between them a mutually satisfactory kin role relationship."[61] Usually they work out something which involves investigating any real kin relationships, e.g. are two women 'related' because one married into and another out of the same *village,* although not the same family? That is, it is based on the fictive kinship relationships described by Freed if they are related to the same village. Sometimes there is fluctuation in a usage--someone gets married after a term has been chosen, or someone didn't know that someone else was related to a certain family, etc. Terms may change if new factors are introduced; but some may not change in order to avoid certain role obligations. Finally, when no prior fictive relationships exist, residents are free to structure their own.

In this they may choose the *sasurāl* or *pīhar*, but usually the women make the choice, and they choose *pīhar*, since it makes them sisters and things are easier--cognates can be more 'loving' than affines. However, this kind of kinship implies that the husbands are in a relationship of *jījā/sālī* to the neighbor women, which implies joking and physical contact. "Those who use these terms. . . are careful to eschew the informal behavior which they imply."[62] Others avoid even this, with the men calling each other brothers, and the women sisters.

Two contradictory assertions arise here. One is that the existence of fictive kinship prevents initiation of sexual liaisons; but the other is that the existence of fictive kinship covers up an illicit sexual relationship. The fictive pattern has the advantage of allowing people to ignore status differences between bridge-givers and -takers, brides and daughters, and between consanguines and affines; the woman can live at ease because she is not in the husband's sasural; she can walk in the *mohallā* unveiled, as in the *pīhar*. The reciprocal usage ignores relative age (otherwise important in the Hindi terminology). "These patterns seem to suggest a desire to minimize distinctions of seniority between urban neighbors and to evade the sense of hierarchy which is inevitably present in the village kinship environment."[63]

Vatuk concludes with the question of how widespread this phenomenon might be in urban India, noting that certain westernized people use English terms ("auntie," "uncle") and/or Mr. or Mrs. plus surname. There seems to be a fruitful area here for further research which might show up regional differences.

7.2. *The Ethnography of Speaking*

7.2.0. *Introduction to the Ethnography of Speaking*

Language is so extraordinarily complex that talking about it descriptively invariably entails gross simplification and concentration on some of its facets at the expense of others. Language is a social "entity," used by humans in social, geographical, and temporal contexts for a multitude of purposes. The full description of these contexts is tantamount to a taxonomy of human experience, behavior, and social structure. Faced with the unwieldiness of such tasks, linguists have generally sought to isolate patterns of human verbal behavior from the "non-linguistic" environments in which they are found. Language is thought of as a code facilitating communication between individuals or social groups. It has a discoverable structure of its own, and, at least in theory, is describable without reference to factors outside of itself. This isolation of the linguistic code is not without cost. By taking such a philosophical position one implicitly disclaims generalizations about language which deal with the use of codes as a function of the social, biological, and intellectual condition of human beings.

Other positions are, of course, feasible. The connection among language, its users, and its contexts, has been of interest throughout the history of western and Indian civilization, and considered an integral portion of philosophy. It is possible to examine the appropriateness of linguistic codes or components thereof to the set of concepts or entities represented by the codes (*physis-nomos* controversy), to the speakers using them, (Whorf-Sapir hypothesis, linguistic relativity) and to the species using them (i.e., rationalist views of language, biolinguistics). Linguistic codes can also be studied with regard to their employment in social contexts. With whom does one use a particular code or portion of one? What are the isolatable social factors which covary with the use of given linguistic features? Language can also be studied from the perspective of its use in manifesting internal states of its users (i.e. the expressive function of language) and in achieving particular ends (pragmatics). Until relatively recently most linguists operated as though the above mentioned aspects of language were describable only after one has achieved a fully satisfactory description of the central facet of language, the code itself. Other aspects of it are paralinguistic and their study accorded a "hyphenated" status (i.e. bio-linguistics, anthropological linguistics, socio-linguistics, etc.). Recent developments in linguistic theory have made it clear that the isolation of homogeneous linguistic codes apart from their contextual use involves a needless falsification of the structure of the codes, and that systematic aspects of linguistic codes covary with various types of "paralinguistic" functions. Many linguists now assert that many types of variability in language use are directly correlated with social, pragmatic, and expressive aspects of the context of language use.

In the first portion of this chapter we explored situations in which some sections of the lexicon of South Asian languages display structures which are susceptible to description. In other words, we dealt with structured sets of linguistic components in which it is possible to demonstrate paradigmatic relations among lexical items. The description of such sets of lexical items has generally centered around kinship terms, color words, vocational or caste terminology, and morphologically manifest categories of verbal and nominal inflection. These studies have more often than not merely pointed out the existence of the structured set of items. There has been a sort of unstated Whorfian assumption to the effect that the presence of a set of structured lexical items is a necessary reflection of a well motivated set of social distinctions. Language serves a mediating function between the hidden cognitive structure of a people and the outside world. Structured sets of lexical items are a handy way of "getting a handle" on hidden cognitive structures. The danger in this line of approach is that the connection between structured lexical items and social structure is not always a direct one, if it exists at all.[64] It grossly underestimates the ability of speakers and linguistic communities to manipulate the ready made linguistic elements which are handed down to them by their cultures. It is necessary to distinguish between the presence of linguistic forms in a code and the employment of those forms. In this chapter we have had the term "ethnographic semantics" to refer to the study of structured sets of lexical items in a language which are thought to embody a meaningful social distinction in the culture using the language. Clearly we believe such studies in and of themselves to be potentially misleading and superficial. We use the term "ethnography of speaking" to refer to the study of the use of language in defined social contexts, for particular purposes, and revealing significant aspects about the social and/or dispositional states of its users.

The ethnography of speaking is a considerably more complex field than ethnographic semantics as we have defined the terms. The latter designates a study whose topic involves interrelationships among the items in the theoretical lexicon of the language. It deals, in effect, with items *potentially* available to speakers of a language. The ethnography of speaking necessarily entails the results of such a study, but adds to its domain all of the factors which influence the *utilization* of any or all of the items potentially available. This can include factors consciously known to the speaker as well as subconscious factors of his culture or environment which are correlated with his selection of linguistic features. The ethnography of speaking can have a number of distinct facets involving language use.

7.2.1. *Aspects of the Ethnography of Speaking*

7.2.1.1. *Speech Acts and Performatives*

The exclusion from most linguistic description of all factors outside of the linguistic code itself has lead to a concentration on the use of language in what has been called by many its "referential function," (i.e. the use of strings of linguistic forms to express propositional formulations). Thoughtful linguists have noted that languages employ linguistic constructions in other functions. Language can be conceptualized of as revealing (albeit not necessarily intentionally) aspects of the emotional state of individuals. In this *expressive* function of language utterances are symptomatic of aspects of their speaker. Language can also be thought of as being intended to elicit a response from listeners. Asking questions, giving commands, etc. are examples of this *conative* function of language.[65] Unfortunately much modern linguistics, particularly American Bloomfieldian linguistics, and transformational grammar until quite recently has relegated the non-referential functions of language to a secondary position in linguistic description. The topic has been discussed much more extensively within the Prague School and other European models of structuralism and by some philosophers of language, most notably J. L. Austin and John Searle.[66]

A theory which subsumes these three distinct functions for the language is possible if utterances are not thought of as sentences but rather as "speech acts." A speech act is performed in order to carry out any of the distinct recognized functions of language. The selection of such a function will often have correlates in the purely formal structure of utterances, or it may be covert. Thus in English the conative function of procuring information from an addressed party may be overtly manifest by the use of a question *morph (what, why, where* etc.), by rising clause terminal pitch without an overt question word (e.g. *He's really selling his house and moving to Australia?*), or it may not be overtly marked at all. (e.g. *John wants me to ask you if you like to play poker)* This last sentence is in a form of a declarative sentence which is being used to elicit information. Speech act analysis of natural languages originally began as a branch of the philosophy of language but has increasingly been incorporated into modern linguistic theory. This has been enhanced by the recognition that many formal aspects of linguistic codes are explicable only in terms of the function they are serving as speech acts.

Tied in to the discussion of speech acts is the description of linguistic "performatives", defined by J. L. Austin as sentences in which "the issuing of the utterance is the performing of action. . ."[67] When one says such things as "I name this ship the *Queen Elizabeth*" (Austin 1961:222) or "We wish you a merry Christmas" (example ours) one is not reporting a situation, but actually performing the actions of naming or wishing. In many cultures the utterances, hymns, chants, aphorisms, or even otherwise meaningless sequences of phonological elements can have performative power, and the analysis of such constructions in purely referential terms is likely to be unproductive.

7.2.1.2. Linguistic pragmatics

The term "pragmatics" is here used to refer to the study of the utilization of formal linguistic devices to achieve (or to avoid achieving) describable consequences. It subsumes a wealth of phenomena such as the use of fictive kinship terminology, circumlocution and euphemism, taboo, no-naming (cf. *7.2.2.3*), and various orders of poetic devices. It also incroaches upon the stylistic manipulation of registers of languages for particular purposes. It can also deal with the conscious manipulation of alternate codes of expression open to an individual or social group (code switching). Virtually any level of linguistic structure can be involved in the pragmatic use of language, and the alternation among entire linguistic codes can also be manipulated. The selection of options offered by linguistic variables for pragmatic purposes can be carried out with conscious knowledge of the purposes of a particular linguistic selection, or can take place without such conscious knowledge.

7.2.1.3. Sociolinguistic Extensions of the Ethnography of Speaking

The study of the use of language in its social context involves aspects of and can have implications for a number of areas of sociology, anthropology, and psychology. The ethnography of speaking is essentially a study which deals with systematic variability in the use of linguistic features. It assumes that some variable features of linguistic structure are not explicable purely in terms of the "linguistic" context. Within much of twentieth century linguistics such phenomena had to be thought of as cases of "stylistic variation" or "free variation". In the ethnography of speaking we expand the class of phenomena which can be allowed to covary with linguistic features. We can point out that the tendency to use [p] rather than [f] is a function of socio-economic class, or that the employment of Sanskritic lexical items in Gujarati instead of native or Perso-Arabic forms is a function of the amount of education or sex of the speaker. We might also show that a bilingual Yiddish-English speaker uses each of those two codes in a set of defined social contexts. But these studies can be taken further. We can demonstrate not only systematic patterns of linguistic variation in the objective patterns of language use, but also structured patterns in the *subjective* reactions of language users to differing realizations of linguistic variables. The fact that a given social group uses one linguistic form more frequently than another by no means indicates that the former group feels the linguistic forms it favors to be more prestigious than another. The study of subjective attitudes towards language use is clearly an important extension of the ethnography of speaking. The correlation of these subjective attitudes with aspects of the sociocultural background of their holders is another important aspect of systematic language use.

It has been implicit so far in our discussion that aspects of social structure can serve as influencing contexts for the exercising of options in the use of linguistic variables. It is possible, although by no means necessarily true, that systematic variation in linguistic variation necessarily reveals structured aspects of social structure. For instance, languages often have two or more second person pronouns, with the use of particular of these forms restricted to certain classes of individuals. The description of such pronominal phenomena can be used to defend assertions of sociological distinctions within the culture. It should be pointed out here that we feel such a line of approach in general to be an extremely dangerous one. Generalizations about social structure as revealed through language clearly need to take into consideration not merely the existence within a language of certain linguistic forms, but also the systematic interrelationship existing between all of these forms (i.e. the paradigmatic relationships between them) as well as observed patterns in the use of the forms. It is unfortunate that many anthropological studies within ethnographic semantics have fallen prey to making generalizations about social structure on the basis of the mere existence within a language's lexicon of a set of structured items.

The discussion of the use of particular codes or components thereof for particular purposes can easily extend to the study of standard languages, and the use of languages for various official purposes. This, of course, enters into the larger question of the politics of language use and of language planning. Suffice it here to note that the according of a particular code standard language status involves in part pragmatic aspects of language use. The "standard" status of a linguistic code may be the natural result of a large number of prior pragmatic decisions concerning the use of this code for socially prestigious purposes. It may, on the other hand, be a reflection of political decision making having little to do with pragmatic value of the code in purely linguistic terms.

The maintenance of a code or reinstitution of one where it previously may have been lost also relates to the ethnography of speaking. A group uses a particular code because it is of some worth to them in achieving some social strategy, or affords them a degree of social cohesion which they may feel necessary for overcoming threats to their group integrity. In some cases the total autonomy of the alternating codes may be abridged and isolated forms or constructions become symbolically equivalent to the use of one of the full codes. The use of particular strata of "multilingual" expanded linguistic competences frequently has subjective value both to its users and to its hearers. The set of associated subjective values of different strata of multilingual competences provides a matrix in which the artful language user can play upon and manipulate elements and levels of language with regard to these subjective values. The ethnography of speaking thus includes the description of the linguistic options open to social groups in sociologically complex situations, the factors which determine how these options are exercised, the subjective values which each of these options may hold for different segments of the social group, the correlations, if there are any, between the objective patterns in the linguistic realizations of options and the subjective attitudes of language users towards these options, and the manipulation of linguistic variability for practical purposes.

7.2.2. *Literature on the Ethnography of Speaking in South Asia*

Unfortunately, the student of the ethnography of speaking in South Asia has limited published material which he can consult to aid him in his pursuits. Until quite recently it was necessary to ferret information from grammars, articles, and anthropological studies whose main purposes were other than the study of language use in its social context. Those studies which have existed have generally been limited in applicability to our topic here, and tended to focus on pronominal usage and terms of address. There is, of course, a substantial literature on kinship terminology and kinship systems in South Asia (discussed earlier in this chapter), but the major portion of this literature is either marginally or not at all relevant for the ethnography of speaking. Of greater importance for us here is an article and a recent doctoral dissertation by Dhanesh Jain, which treats the verbalization of respect in Hindi. Jain, a student of Dell Hymes and William Labov, is probably the first scholar to attempt to incorporate forms of address, taboo, circumlocution, and other linguistic devices into a cohesive theory of how language is used by its speakers to accord degrees of respect in a South Asian language. Other than Jain's and a few other studies, it is obvious that few sociolinguistically sophisticated studies of the ethnography of speaking in South Asia have as yet been carried out. In the following sections we attempt to examine the small number of studies which are linguistically relevant to the subject, to point out some of their limitations, and outline fruitful areas of investigation in the future.

7.2.2.1. Terms of Address and Pronominal Usage

It is a characteristic feature of South Asian languages to have multiple pronominal forms which are available for a speaker in either addressing or referring to an individual or individuals. Hindi, for instance, has three second person pronouns, *tū*, *tum*, and *āp*, all of which can be used in addressing or referring to single individuals, and of which the latter two can be used in addressing groups. Many languages have multiple third person pronominal forms which can be used in referring to individuals. The set of pronouns in Hindi can be structured into a system in which person and number are independent dimensions, and in which they determine grammatical features such as verbal agreement and adjectival (i.e. participial) agreement. The Hindi non-oblique forms, (i.e. forms used generally when preceding a postposition) thus form a system such as the following:

	Singular	*Plural*
1st person	mãī	ham
2nd person	tū	tum, āp
3rd person	yah (proximate)	ye (proximate)
	vah (non-proximate)	ve (non-proximate)

Some grammatical features of Hindi sentences are determined by the place of an employed pronominal form in the above chart (e.g. which form of the verb is used with each pronoun.) All competent grammars of Hindi provide this information. They generally, however, only provide perfunctory information about the choice of a pronoun from among competing forms. For example, it is common in many Hindi vernacular dialects to use the plural first person pronoun *ham* in addressing oneself. This usage alternates with the more standard form *mãī*. Grammars tend to dismiss the plural usage as substandard, uneducated or vulgar. We would like to know under what conditions the usage is grammatical. By whom, with whom, and for what purpose may it be used? What does its use reveal about the speaker? What do Hindi speakers feel about such usage? Do they consider it substandard? Similarly, which second person pronouns are used for what persons by what speakers? Is it the case, as has been implied by certain linguistic descriptions, that it is an inherent quality of an individual that he be addressed by one and only one of these three second pronominal forms? Is the pronoun used to address him a function of his genetic and social position in north Indian culture? Pronominal systems parallel to the Hindi system exist for all South Asian languages, and all have socially determined options in the selection from among alternate pronominal forms.

7.2.2.1.1. A. Chandrasekhar (1970) gives an inventory of pronominal forms available for use in Malayalam, and a cursory description of by and for whom each of the forms is used. Pronouns are broken up by person and an indication provided of by and for whom each is used. In the case of second and third persons there can be numerous choices open to the speaker and it would appear that there is much overlapping in the range of use for many of the forms.[68] Chandrasekhar also provides a summary of the first and second person pronominal forms used in a dyadic relationship. Each dyad is conceptualized as having a sender and a receiver and information is provided on the form used by the sender to refer to himself, the form used by the sender to refer to the receiver, the form used by the receiver to refer to the sender, and social characteristics of the sender and the receiver.

The Chandrasekhar article also cursorily mentions some related phenomena of pronominal usage which are of linguistic interest. The selection of particular second person pronouns requires the selection of specific nominal forms elsewhere in the same sentence. The choice of the second person pronoun *tiruměni* or *tirumanassə* requires the use of a special set of forms referring to body parts formed with the prefix *tiru (tirumukham* 'the auspicious face', *tiruvayarə* 'the auspicious belly', or other specialized forms in place of more common ones (*nīrāṭal* instead of common *kuḷi* 'bath'). Similarly the selection of a non-respectful first person pronoun will lead to the use of the humble forms *kuppāṭə* 'dirt heap' for *vīṭə* 'house'. (Chandrasekhar 1970:250)

Unfortunately Chandrasekhar does not provide us with a sophisticated set of sociological factors to condition the selection of the alternate pronominal forms which he describes. He does divide the third person forms into five categories depending on "the social status of the referee as well as the social status and mental attitude of the speaker." (Chandrasekhar 1970:250) Factors cited by Chandrasekhar in illustrating these forms include superior vs. inferior social position, position on a hierarchical scale of castes, amount of respect accorded an individual, equality of status between members of dyad, insult, royal status, friendship, degree of formality, education. These social factors are not, however, developed into a coherent system.

7.2.2.1.2. A description of a similar, although morphologically simpler system in Bengali is given in Das (1968). Das discusses second and third person pronominal forms and then attempts to describe the range of forms which can be used to address individuals. The use of the three second person pronominal forms *apni*, *tumi*, and *tui* is described in terms of a family situation involving eight individuals.[69] Das summarizes the use of particular linguistic forms by members of dyads drawn from these eight parties in the chart reproduced below.

Addressor	*Addressed*	*Form of Address*	
F, M	N	*apni*	↔
B, Eb, Yb	N		→
S	N, F, M		→
Fr	F, M		→
F, M, N	own children	*tumi*	↔
	other's children		→
F	M		↔
B, Eb, Yb, Fr	(S)		→
B	Eb		↔
Eb	(B, Yb) Fr		↔
Yb	B, Eb, Fr		↔
F, M, N	own children, S	*tui*	→
B	Fr		↔
Eb	B, Yb, (Fr)		→
everybody	S		→

Symbols within parentheses in the chart mean that the person 'may be addressed'. ↔ means reciprocal. → means only used by the addressor. (Das 1968:21)

Figure 62. ***Terms of Address in Bengali***

Das's description of non-pronominal forms of address is in many ways more interesting than that of the pronominal forms. He formulates a series of rules governing the use of names, and their cooccurence with what he calls "address words" (e.g. English *Mr.*, *Sir*, Bengali *bābū*, *mɔsai*, *saheb*, etc.). He discusses the correlation between the use of various address words in conjunction with portions of all of an individual's proper name with the use of different pronominal forms.

Das also includes in his article a discussion of Bengali kinship terminology. He points out that there are at least two distinct functions for such terminology, referring to an individual and addressing him. He groups kinship terms into those which have identical forms for the two functions and those which don't (*kākā* 'father's younger brother' is used for both functions but the individual referred to as *sosur* is addressed as *baba* in conjunction with address words when appropriate).

7.2.2.1.3. Karunatillake and Suseendirarajah (1975) have attempted a comparative study of the use of second person pronominal forms in Sri Lanka Tamil[70] and Sinhalese. The stated purpose of their article is "to investigate the several social differences that are reflected in the language spoken by the different communities." (1975:84) Karunatillake and Suseendirarajah maintain that differences in the patterns of pronominal usage between the two communities are correlated with "the social differences in each community mentioned. . ." (1975:84) They also make the stronger, and in our opinion unsupportable claim, that "it is possible to describe the social hierarchy in our [Sri Lankan] society on the basis of the use of pronouns of address." (1975:88)

In describing the use of Tamil second person pronouns *nii*, *niir*, and *niimkaḷ*, Karunatillake and Suseendirarajah emphasize the importance of a number of social aspects (caste, age, education, position or rank, sex, wealth, family background, personality) of the person addressed, seeing these social factors subsumed under a more general notion of "social status". The absence of such status falls together with the notion of "intimacy between participants of the speech event." Roughly speaking, social status of the addressee is indicated by the use of *niimkaḷ*, absence of respect or social intimacy by *nii* and 'medio-intimacy' by *niir*. They give illustrations of the use of these forms in these senses in a number of types of social situations (husband-wife, mixed caste and age, etc.). They also describe the use of a more restricted two-form second person pronominal system (*nii* and *niimkaḷ*) among Sri Lanka Tamil speaking Muslims (which they say parallels the use in Indian Tamil). The least acceptable aspect of Karunatillake and Suseendirarajah's paper is their assertion to the effect that it is possible to make a three-way division of Sri Lanka Tamil society based on the use of second person pronouns: "one group deserving the use of *nii* alone; another group the use of *niir* and the third group *niimkaḷ*--from the point of a high-caste, educated, and aged person." (1975:88) The determining factors governing the use of forms in the Sinhalese second person pronominal system are fundamentally different from those at work in the Sri Lanka Tamil system. Karunatillake and Suseendirarajah consider the primary binary division in Sinhalese society to be between the clergy and the laity. They describe forms usable in contact between the two groups.[71] Within the clergy, the major divisions are between teachers and pupils and novices and non-novices. In addition, the relative status within the clerical hierarchy of the participants of the speech event is a determining factor of pronominal selection.[72] Among the laity, the selection from among six pronominal forms of address *(eyaa, tamuse, um̆bə, too, baṅ, bolə)* is determined by "age group, sex, social status, external appearance, and personality, etc."[73] Karunatillake and Suseendirarajah cite a number of linguistic phenomena which covary with the selection of specific personal pronouns. Included here are verbal command forms and consentual positive response words to direct yes-no questions.[74] Although Karunatillake and Suseendirarajah consider both Sri Lanka Tamil and Sinhalese society to be compartmentalized in ways paralleled by pronominal address usage, they maintain that the criteria used in determining the correct address forms in the two cultures are fundamentally different:

> The governing factor in the Tamil society being social, primarily caste and social-rank based, and as such the choice of the appropriate pronoun being already made for the speaker; whereas the selection of an addressee-pronoun in Sinhalese depends more on the attitude of the speaker towards the addressee and seems to be more a psychological fact than a socially preconditioned fact. Although there is so to say a semi-functional caste system underlying the Sinhalese society it is seldom realized at the level of speech; and, therefore, unlike in the Tamil society, caste and social-rank distinction never function as a determiner of the addressee pronoun-selection in Sinhalese. What is crucial for the Sinhalese society is the ranking of a person along a politeness (respect) scale, whereas

> in Tamil it is a socio-economic scale. In Tamil, the ranking of a person is rather rigid as revealed by the specified linguistic exponents, whereas in Sinhalese it is more flexible. In the Sinhalese society age can be considered as one of the constant factors conditioning the selection of an appropriate pronominal along the politeness scale, whereas in the Tamil society other factors such as social rank and caste seem to be dominant. (Karunatillake and Suseendirarajah 1975:95)

Once again it should be pointed out that conclusions about social structure based on linguistic evidence must be looked upon with caution. Karunatillake and Suseendirarajah have demonstrated that the selection of particular pronominal forms of address in Sri Lanka Tamil and Sinhalese seems to be correlated with social properties. It remains to be seen to what extent these social variables are equally well-motivated in terms of covariance with other Sri Lankan social phenomena. Furthermore, we have little evidence about whether the forms they describe were reported by native speakers as being used in particular contexts, or were actually observed by impartial investigations. It is essential that we distinguish in investigations such as these between objective patterns of linguistic variation and both subjective impressions of these patterns and the attitudes of native speakers towards the use of socially conditioned linguistic alternatives. Objective patterns of linguistic behavior cannot be ascertained without carrying on extensive data collection conducted with scientifically selected sets of informants. Chandrasekhar's Malayalam and Das's Bengali data are subject to exactly the same cautionary notes.

7.2.2.2. Functional Usages of Address Forms and Pronouns

7.2.2.2.1. A sophisticated critique of purely cognitive models of kinship terminology is found in Tyler's study of the use of kinship terminology in Koya (Tyler 1968). Tyler strongly criticizes descriptions of kinship terminology which are carried out solely in terms of the "genealogical denotata" of the nominal stems of kinship terms. He argues that adequate descriptions of the use of these forms requires increased emphasis on the social and linguistic contexts of these terms. Tyler derives his data from a number of villages along the banks of the Godavary near the road from Bhadrachallan to Nugur. The Koya informants whom Tyler worked with were frequently in contact with Telugu speakers, and some members of the community were bilingual Koya-Telugu.[75] Tyler provides an analysis of both Telugu and Koya traditional kinship terminology, and adds a componential analysis of the semantic structures of the two codes. Independent parameters of these two cognitive systems are (1) sex, (2) generation, (3) cross vs. parallel, and (4) relative age. (Tyler 1968:256) He constructs rules which account for minor differences in the componential structure of the two kinship systems. Once these conversions are accounted for, Tyler allows for the change from a kinship term in one of the languages to a corresponding one in the other by calqueing (translating the morphemes of a word of one language to the corresponding morphemes of another.) In addition, Tyler points out sets of lexical items in both languages which refer to identical sets of componential features. In the Telugu system, for instance, the terms for father (*ayya, taNdri, nayana, naana*), mother *(aama, talli)*, son *(koDuku, kummaraDu, abbayi)*, younger sister *(cellelu, celli)*, and daughter *(ammayi, kuturu, biDDA)* as well as the terms for father *(eyya, tappe)*, mother *(evva, talluru)*, son *(marri, peeka)*, and daughter *(mayyaaDi, piikiDi)* in Koya have multiple realizations. He notes that some of this alternation among speech forms can be explained in terms of the formality of the situation of the speech event. Thus *taNdri* can be distinguished from *nayana* in situations of greater formality.

Although Tyler is to be commended for urging greater attention to the social context of the use of kinship terminology, his analysis of the contexts of Koya and Telugu terms is only in terms of such crude factors as formal vs. informal, intimate vs. non-intimate setting, in the home vs. outside of it, etc. No attempt is made to formulate a systematic theory of social contexts of linguistic alternation, or to set up criteria by which one can identify a specific social variable.

7.2.2.2.2. Vatuk in two papers (1969a,b) tries to explore how Hindi speakers are able to manipulate abstract genealogical kinship systems known to them in finding terms to address individuals in complex and changing social environments. In one of these papers (1969a) she discusses reference and address in the new *mohallās* of a newly settled medium-sized city in Western Uttar Pradesh. The situation is of particular interest to her because of the inclusion of occurrences of transitory social interaction which would have been unlikely to occur in rural villages or older more stable *mohallās*. Vatuk demonstrates that in seeking acceptable forms of address for use in new social settings, Hindi speakers are able to cull from a wide variety of formal linguistic devices, and, more significantly, that the pattern of linguistic usage in such circumstances does not correspond directly to the kinship or fictive kinship systems used in home villages.[76] She argues that traditional Hindi kinship provides a large stock of terms which is available for exploitation in referring to and addressing individuals. The selection of forms from the options available from traditional cognitive systems "reveal[s] much about the traditional social structure and the direction of change under urbanization."[77] She also demonstrates that the use of kinship terminology is only one of a number of formal devices which are used for referential or addressing functions. She outlines a number of these devices, stating wherever possible the linguistic correlates and the social context of their use.

According to Vatuk there are at least six distinct modes of naming used in the *mohallā* of her investigation, in addition to the use of an individual's surname, given name, or both, either with or without such forms as *shrīmati* 'Mrs.", *Mr., Mrs.* [and we assume *kumārī* 'Miss', *shrī* 'Mr.', etc.]. Vatuk points out that names are used only in limited contexts.[78] The other six types of address processes are outlined below:

> 1. *Caste:* (a) the masculine or feminine form of the caste name itself, followed for high-caste persons by the respect suffix *-jī*, e.g. Tyāgī/Tyāginī, Nāī/Nāīn; (b) a caste "surname", e.g., the name of a subcaste or gotra or some other name associated with a particular caste, with the respect suffix *-jī* or *sāhab*, for men only, e.g. Sharmājī [Brahman], Guptājī [Baniā], Goyal Sāhab [Baniā]; (c) a caste title, with the respect suffix, e.g. Paṇḍitji/Paṇḍitānī [Brahman], Lālāji [Baniā], Chaudhurī [Jāt]; (d) a caste name with the genitive postposition *-kā, -ki, -ke,* "[child] of," e.g., Brahman-*ki* [Brahman girl], Nāī-*kā* Nāī boy, for children and persons of low social status.
>
> 2. *Occupation:* former occupation, academic degree (with or without the respect suffix); for woman, occupation or title of husband, e.g., Patvārī/Patvārni [village official], Vakīl Sāhab/Vakīlnī [lawyer], Tekedār/Tekedārnī [contractor], Māstarjī/Māstarni [school teacher], Doktor Sāhab/Doktornī [doctor], Stationmaster, Postmaster, Shāstrījī/Shāstarni [holder of Sanskrit degree], Bahenjī [lady school teacher, lit. "sister"], Naukarānī [servant woman].
>
> 3. *Place of origin:* with suffix *-vāle, -vālī,* "man/woman of," (e.g., Hapurvāle, Dillīvāli).

> 4. *Teknonymy*: especially for women, e.g. Rāmū-*kī mā* [Ram's mother], Omī-*kī bahū* [Om's wife].
>
> 5. *General terms:* referring to common *mohallā* residence, generally prefixed by *hamāre, -ī*, "our", e.g., *makānmālik/-in* [landlord/-lady], *barābarvāle/-ī* [next-door neighbor], *parosī/-in* [neighbor], *kirayādār* [tenant], *mohallevāle/-ī* [resident of the *mohallā]*.
>
> 6. *Kinship terms:* used in reference only when the person spoken to can discern clearly from the context the object of reference. A clarifying pronoun *(tumhāre*, "your," *mere*, "my") or name and genitive postposition may be prefixed. . . (Vatuk 1969a:256-7)

The fictive use of kinship forms is particularly widespread in Vatuk's *mohallā*. She states that it is the most commonly made use of address by women and children in addressing *mohallā* neighbors, and is also frequently used by men. She argues that the true kinship system is the predominant form of address used within the relatively closed social systems in the villages of origin. The overwhelming majority of individuals with whom people, particularly women and children, come in contact in village life are referrable to within the kinship system. Family relocation to a new *mohallā* involves the establishment of a network of social relationships largely unknown in the village. The traditional kinship system serves as a template against which individuals can establish new social relationships, and, equally importantly, act with individuals in ways appropriate to their age, caste, education, occupation, sex, and overall social position. The use of fictive kinship terms, of course, does not completely overshadow the use of terms indicating occupation, caste, place of origin, etc., but does allow the establishment of a sufficient number of social links among *mohallā* residents to create a metaphorical extension of the village family structure. Vatuk points out that when *mohallā* residents meet for the first time there is a strong tendency to establish kinship links as quickly as possible, and thus eliminate uncertainty and discomfort in address. She states that there are two major factors which can be used in determining the appropriate form of address between the neighbors: 1. "childhood residence of Ego or his spouse in the present *mohallā* or an adjoining *mohallā*, and 2. the existence of real or "village kinship ties between Ego or his spouse, and the *mohallā* residents." (Vatuk 1969a:263) In other words, an immediate attempt is made to establish connection via the traditional kinship system.[79] If this fails, the residents "are free to structure their own fictive kinship with neighbors as they become acquainted."[80]

7.2.2.2.3. A completely different aspect of address and reference is discussed in Jain 1969. Jain is primarily interested in the verbal aspects of according (or not according) respect both in addressing individuals or referring to them. He states that "respect is an important feature of the social action of Hindi speakers,"[81] meaning that the degree of respect which one wishes to accord an individual is a conditioning factor in the selection of a proper mode of address. He correctly observes that appropriate forms of reference and address are only one type of behavior associated with the manifestation of respect, and that their use parallels non-verbal actions such as "greeting someone with a bow, with raised or folded hands, sitting on a lower platform than that of the recipient of respect. . ."[82] Jain finds that the selection of proper forms of respect in Hindi is governed by "the participants, their relationship to each other, the social situation and certain other factors."[83] He explores the linguistic phenomena which are the manifestations of respect, as well as attempts to describe the social phenomena which trigger their use.

According to Jain there are several linguistic phenomena in Hindi which are used to manifest degrees of respect. Some of the most important of these are (1) the selection of an appropriate combination of name plus honorific word or title;[84] (2) the selection of vocabulary by a speaker indicating the insignificance of himself and his possessions and the high quality of the addressed and his possessions,[85] (3) the selection of foreign rather than native respect forms,[86] (4) the use of grammatically plural forms, both in address and in reference, with the selection of such forms demanding the use of plural grammatical agreement features (i.e. in verbal agreement with plural nominals, with adjectives modifying plural nouns, in the selection of appropriate command forms, etc.),[87] (5) use of appropriate 2nd person pronominal form in address situations,[88] and (6) the avoidance of addressing or referring to the name of an individual to be accorded respect.[89]

An important aspect of the Jain article is its demonstration that the decision to accord respect in Hindi is not purely a function of the inherent social position of the person addressed, but also involves the nature of the addressor, the context of the linguistic event, the reciprocity or lack of it between the speaker and addressee, and the intention of the speaker. Jain observes that reciprocity of the use of linguistic respect devices is a function of the social distance between the members of the dyad. Where the social distance is minimal, the use of forms of address is likely to be reciprocal, as in the case of students or friends. The choice of the specific pronominal form used by parties who share a reciprocal respect relationship is likely to be determined by the context of the speech event and the amount of time the parties have known each other.[90] Jain also correctly observes that the selection of the proper pronoun of address is influenced by the intention of the speaker. The total set of social relations holding between speaker and addressee contribute to an *expectation* of the appropriate form of address. This expectation need not be fulfilled, and may, in fact, be contradicted for any of a number of purposes.[91] Another interesting point made by Jain is that violation of rules of appropriate forms of address and reference leads to the production of what are, in at least one sense, ungrammatical sentences. This notion of grammaticality, however, has to be understood as indicating sentences which are inappropriate for utterance by a particular individual within a prescribed context and for a specific purpose. No sentence involving reference or address can thus be thought of as inherently grammatical or ungrammatical, but is rather appropriate or inappropriate to the specific context.[92]

7.2.2.2.4. An additional attempt to explore the verbalization of respect in a South Asian language is contained in Suseendirarajah (1970), which purports to "portray briefly how some social aspects of modern society of the Jaffna Tamils are reflected in the language they speak." (Suseendirarajah 1970:239) The paper is essentially an inventory of linguistic features which can be associated with differences in social status between the members of a speech event (or between a participant of a speech event and a person referred to) or with different amounts of respect accorded someone in a speech act. The features listed include different forms of the imperative indicating differential degrees of respect;[93] different morphemes used in the second and third person finite verb forms depending on the social status of the individual referred to;[94] variability in the use of the second person pronouns *nii*, *niir*, and *niiṅkaḷ*;[95] the use of a nominal suffix *aaḷ* designating medial respect;[96] variability in the use or non-use of caste names as terms of address;[97] the use of different terminations for proper names, occupational terms, and kinship forms to indicate varying degrees of respect.[98] Suseendirarajah sees a generational split among modern Jaffna Tamil speakers, with the younger generation demonstrating a willingness to give up some or all of the societal distinctions in their speech.[99]

7.2.2.3. An extreme extension of the notion of "appropriateness of a linguistic item to a specific context" is taboo, in which some form of behavior is proscribed. The forbidden behavior is very frequently verbal, and the prohibition may take any of a number of forms. A given object or concept may be taboo, and the word or words representing it may also become objectionable. In some cases a given sequence of phonological elements (as apart from the semantics associated with those elements) may become avoided, frequently because it is associated with the phonological sequence representing a tabooed item or pattern of behavior. In some cases it is not an item or concept which is tabooed by a culture, but only a particular manner of referring to it. Where this is the case a language needs alternate devices for expressing the concept. Many types of verbal taboo prompt the creation of circumlocutions, euphemisms, and other linguistic devices, allowing the bypass of some unacceptable means of expression.

The general literature on taboo is vast, and we are unable to examine any significant portion of it here. Most of this literature focuses on religious and/or cultural aspects of taboo and on how the items, concepts, etc. which are proscribed fit into overall patterns of religious belief, social behavior, etc. Here, however, we are interested in only the linguistic correlates of taboo, particularly with regard to how they effect models of address and reference. It is clear that verbal notions of taboo in South Asia are closely connected with the verbalization of respect, as described by Jain, and that the avoidance of classes of linguistic forms and constructions is part of a structured system of expressing respect. It is also clear that many linguistic devices in Hindi, and presumably in other South Asian languages, arise from the attempt to find alternate means of expression for concepts whose most direct path of expression is proscribed.

7.2.2.3.1. Bharati (1963) attempts a description and explanation of the avoidance of kinship terminology in referring to individuals who are accorded respect, particularly by a wife in referring to her husband. It has been observed elsewhere[100] that in many South Asian cultures there is a restriction on pronouncing the name of one's husband, and that a number of circumlocutionary devices are resorted to when a woman needs to address her husband. Bharati claims here that such verbal avoidance in Hindi cannot be considered an unqualified taboo, and the language and culture provide alternate modes of expression for the semantics of the proscribed item. He reports that in his area of investigation, that of Baniās communities in Delhi, and selected urban areas of Uttar Pradesh and Bihar, a new pattern has arisen in which English lexical items are allowed to substitute for Hindi ones. The use of sentences with these English forms is felt by Bharati to avoid some of the taboo restrictions which would pertain to corresponding sentences with Hindi items.[101] This substitution also extends to contexts in which one might not expect the substitution of non-native lexical items. He reports such sentences as (spoken by a Delhi Baniā) *parsõ merī sister kī shādī ho rahī hai* where one would expect the use of H. *bahen* for E. *sister*. Bharati finds a rather peculiar correlation between the amount of English known to speakers and the amount substituted in avoidance constructions. He reports that in central Uttar Pradesh individuals with good knowledge of English have a high propensity for using Sanskrit forms, while *baniās*, *kayasths* and other urban dwellers with less extensive English background tend to use liberal doses of English words mixed in with their Hindi. Bharati argues that the substitution of English words for Hindi is carried out by urban middle class professionals and white collar workers, who do not fully want to be associated with orthodox backgrounds, although they by no means wish to consciously reject it. The adoption of English loan items allows the speaker to disassociate himself by his speech patterns with some of the more restrictive aspects of his conservative background, while at the same time permitting him to maintain his public religious identity. Bharati sees the utilization of English forms helpful in the creation of a "subidentity"[103] and as a "psychological relief medium."[104]

7.2.2.3.2. The only systematic treatment of the linguistic consequences of taboo appears in Dhanesh Jain's doctoral dissertation. Jain presents a discussion of what he calls 'no-naming', defined as "the strategy by which Hindi speakers avoid having to say the name (FN) of the addressee or referent."[105] The phenomenon applies primarily in affinal relationships, and is observed most strongly by a *bahū* (wife living in her husband's joint family). Normally the wife will not utter the name of her husband or her husband's father, although it may sometime be expanded to include all of a wife's elder affines. The extent to which no-naming is observed by a wife varies as a function of a number of social variables, education, caste, income, etc., but in general it is safe to say that the forces of modernization and education work against its utilization, and that religion and tradition support it.

Jain discusses several of the formal devices that can be used by a wife to avoid uttering the name of her husband or other parties. Among the most important of these devices are (1) teknonymy (the practice of naming apparent from his or her child). Thus a wife might refer to her husband as *lallū ke bābū*, ('Lallu's father'), using the plural form of the noun *lallū* 'father'. Similarly, a wife might, while talking with her younger sister, refer to her husband by the term *tumhāre jī jī* ('your elder sister's husband'); (2) the use of vocative expressions and other linguistic constructions which are devoid of any pronominal force, or which have no reference to the person addressed (e.g. *e jī* 'hey!', *mā̃ī ne kahā* 'I said [that. . .]', *bhaī* (used to call attention); (3) the use of third person plural pronouns (*ye, ve,* oblique *un* and *in*) in contexts where it is clear who the referee is; (4) the use of a fictive name. Jain reports that a husband will on occasion call out his daughter's name in the hope of drawing his wife's attention.

Within traditional Hindi culture, the taboo on speaking a husband's name may have interesting consequences. With many married wives, the phonological shape of the husband's name may become so strongly tabooed that other words phonetically similar to the husband's name may also be avoided. Jain cites a number of examples of this phenomenon. In one case a wife's husband's elder brother was named *dhani rām*, and the word *dhaniyā* 'coriander' became avoided by extension. In referring to the frequently used spice this woman used the phrase *harī botal wālā masālā* 'the spice in the green bottle'.

As did Bharati before him, Jain notes that the system of no-naming is undergoing rapid change, most likely due to education, urbanization, and the gradual disappearance of *pardā*. He also notes that technology may have an accelerating effect on the loss of no-naming, as avoiding the name of one's own husband on the telephone may be virtually impossible to carry out on some occasions.[106] The practice of no-naming, however, is still widespread, particularly in the more traditional sectors of society, in non-urban settings, and where westernization is not particularly advanced. When observing the custom, women will frequently go to great lengths to avoid uttering the name of her husband or his kin, even if directly pinned down by the curious investigator.[107]

7.2.2.3.3. The framework developed by Jain for describing the expression of respect in Hindi has been adopted for Tamil in a paper by M. Shanmugam Pillai (Shanmugam Pillai 1972). The phenomena described by Shanmugam Pillai for Tamil are essentially the same sort as those reported by Jain. Like Jain, Shanmugam Pillai notes that the selection of an appropriate mode of address or reference between two parties is a function of their individual social status, mutual relationship, and the context. He divides the status between speaker and the referent or the speaker and the address into four types: (1) respect status, where the speaker respects the recipient, but the recipient does not [respect the speaker]; (2) neutral status, where neither the speaker nor the recipient expects or expresses respect; (3) non-respect status, where the speaker uses non-respect terms, while the recipient uses respect terms; (4) disrespect status, where the speaker uses abusive terms and the recipient does not."[108] Shanmugam Pillai also notes the phenomena of no-naming, and provides examples of word substitutions based on phonological similarity of a form to a husband's name (e.g. the substitution of *ravaṇo/lavaṇo* (Skt. 'salt') for T. *uppɨ* because of the latter's perceived resemblance to the proper name *subbayyaa*.[109]

7.2.2.4. All of the works cited in this section have made the stated or implicit assumption that an examination of systematic variability in particular aspects of languages, especially pronominal systems, address conventions, and the formal devices by which respect is accorded, can lead to an understanding of fundamental social distinctions in a culture. Such a view considers that the language of a group necessarily filters its conceptual perception of the universe, and, conversely, will embody those concepts and generalizations which it finds useful. Franklin Southworth, however, in a recent paper questions whether this relationship is necessarily so direct. More particularly, he examines linguistic change in the complex sociolinguistic patterns of language in expressive power and status relationships. It is his contention that in some cases South Asian languages have altered the socially accepted formulae for manifesting power. Southworth sees this not as a fundamental change in the nature of power relationships in the society, but rather as a mask rendering it difficult to sense the true nature of these relationships.

Southworth deals with a number of changing phenomena which in some way or other are indicative of power relationships in South Asian languages. The first of these deals with the use of terms designating members of a formerly untouchable caste in a Tamilnadu village. The traditionally used term for untouchables was *paraiyan,* but the term *harijan* (literally 'born of God'), created by Mahatma Gandhi in an attempt to raise the condition and self-esteem of untouchables, has been gaining currency among former untouchables in referring to themselves, and among non-untouchables who are aware of the fact that their speech patterns are being observed and judged. Southworth claims that although the total stock of lexical items available for referring to this group has changed, the existence of the group as a social entity, as well as the low power and/or prestige status of this group, has been fundamentally unchanged.[110]

Another example of change in the use of power relationships involves Malayalam pronominals in a Kerala village. Southworth describes a case where an educated Nambudiri Brahman man relates how any young person currently [regardless of caste] uses the polite plural second person Malayalam pronoun *niingaḷ* for an older person--one might expect that a high-caste youth would employ a non-respectful second person pronoun in addressing an elder lower caste adult--while the man's 12-year-old daughter was almost simultaneously heard addressing a 50-year-old Nayar woman with the informal singular second person pronoun *nii.*[111]

Southworth also points out a change involving a semantic recategorization of social classes. Of late the term *kriṣikkaaran* (person who does farming) has been adopted by larger landlords in many parts of Kerala. This is an attempt to equate their economic interests with those of some other classes of farm workers, and thus diffuse potential hostility to themselves. Southworth argues that by the large landowners adopting a term which lumps them together with smaller landowners, they are in effect making "an attempt to try and convince the small landowners that demands by the poor are against the better interests of the small self-cultivating holder who might occasionally employ an extra hand."[112]

An additional socially conditioned change in a cognitive system which is described by Southworth involves the internal structure of second person pronouns in South Asian languages. Southworth reports situations in which there has been increasing reluctance to use first person singular pronouns (and verb forms in agreement with them) for individuals who previously were thought of as being socially inferior, or unsuitable for the according of respect. Southworth notes that there have been reports of greater reciprocity in the use of pronominal forms among husbands and wives (particularly among young educated couples living in urban centers).[113] He inquires about the extent to which changes in the usage of these pronominal forms reflect alternations in the fundamental social distinctions of Indian societies. He concludes that they do not, and that they are rather indicative of a change in the "degree of freedom which the employer feels in asserting the difference[s]."[114] It can, of course, be argued that a change in the attitude of speakers towards using terms embodying a repressive set of social distinctions is a necessary step in the eradication of those distinctions. But it is certainly the case that failure to utilize these terms in particular contexts can equally well indicate a heightened awareness of the social distinction. This awareness is as likely to lead to an increase in behavioral patterns based on the social distinction as it is to lead to the elimination of the distinction.

7.2.2.5.0. Conclusions

The previous discussion of current research on the ethnography of speaking on South Asia points to the paucity of hard data available on the subject. There is no aspect of the field for which there is not a critical need of further data and analysis. The data on the ethnography of speaking in South Asia has been virtually restricted to those aspects of it which relate to pronominal systems, address and referential conventions, no-naming, and kinship systems. This is almost certainly because the field has originated from cognitive studies of structured sets of lexical items, and because the relative cohesiveness of these linguistic subsystems has provided ready access for investigators. Yet we have little valuable data even in these areas.

Studies on the ethnography of respect have been carried out for a limited number of languages. Few studies of pronominal, address usage, etc., are available which stress the criteria used by speakers in making selections among available sets of forms.

Perhaps even more significantly, vast areas of the ethnography of speaking in South Asia are totally unexplored. How, for example, may different sentence types (commands, questions, requests, etc.) be employed to elicit different kinds of information. How can one manifest different types of agreement, and what are the cultural correlates of such agreement.[115] All languages have expressions and formulae whose function is other than transmitting the propositional content of the strings of words uttered. Words, phrases, and sentences have a wide range of functions, and are appropriate to a definable set of circumstances. What is appropriate verbal behavior in a society and what is not? How does one go about asking someone their name, age, occupation, caste, hobbies, etc.? Grammatical analysis of South Asian languages alone cannot answer these questions for us. Data will need to be collected and analyzed for all languages of the subcontinent.

The data collection is a particularly important one for the ethnography of speaking in South Asia. We have learned in the last fifteen years that investigations of the social use of language need to distinguish between observed patterns of linguistic variation, conscious impressions of such alternations, facets of social structure which covary with linguistic variation, subjective impressions about the ways in which facets of social structure covary with linguistic variation. Thus we need to keep distinct statements to the fact that (1) the use of Sanskritic consonant clusters alternates with the simplification of these clusters in Sanskritic *tatsama* vocabulary; (2) Hindus think the use of Sanskritic consonant clusters covaries directly with sex, education, and religion; and (4) educated people believe that Hindus use more Sanskritic consonant clusters than do non-Hindus [but uneducated individuals are less inclined to believe it]. Propositions of each of these four sorts needs confirmation by different means. Labov has shown clearly that distinct data sampling techniques need to be created for eliciting information about language use in different contexts, in differing degrees of formality, and in which the speaker has differing levels of awareness of the observation of his own speech. All too often discussions of the ethnography of speaking in South Asian languages have relied on off-the-cuff observations with little systematic data base. Moreover, we have observed few applications of many contemporary models of sociolinguistic variation to South Asian situations. We hope that the next decade will partially alleviate the paucity of research in this area.

NOTES

1. Hymes, 1962.

2. Hymes, 1962:131.

3. Hymes, 1962:132.

4. Southworth and Daswani, 1974:192-3.

5. Lounsbury, 1964:205.

6. Ibid.

7. Wallace and Atkins, 1960:367-8.

8. Ibid.

9. Frake, 1962:829.

10. Ibid., 30.

11. Ibid., 38.

12. Ibid., 38-9.

13. Ibid., 39.

14. Ibid.

15. Tyler, 1969:3.

16. Ibid.

17. Ibid., 4.

18. Ibid., 5.

19. Ibid., 6.

20. Ibid., 7.

21. Ibid., 11-12.

22. Ibid.

23. Yalman, 1967:209.

24. Ibid.

25. Ibid., 210.

26. Ibid., 213.

27. Ibid.

28. Yalman, 1969:625.

29. Ibid.

30. Tambiah, 1965:168.

31. Ibid., 169.

32. Emeneau, 1938b:337.

33. "The kinship term *guŋgu* (22) may be cited as an example. In Mundari *guŋgu* means great grandfather and great grandmother, great granduncle and great grandaunt, father's brother and his wife, mother's elder sister and her husband, and also the corresponding relations for all of the above. But in Santali which, being spoken in the plains, is more under the influence of Indo-Aryan society, the meaning of the term (i.e. its Santali cognate) has now been restricted to 'father's elder brother' and reciprocally 'younger brother's children.'" (Bhattacharya 1970:462)

34. Yalman, 1962:556.

35. Ibid., 563.

36. Ibid., 567, figure 4.

37. Ibid., 569.

38. Yalman, 1962:567, figure 4.

39. Tyler, 1965:1436.

40. Ibid.

41. Ibid., 1437.

42. Ibid., 1438.

43. Tyler, 1966:489.

44. Ibid., 490.

45. Ibid.

46. Ibid., 493.

47. Ibid., 494.

48. Ibid., 500-1.

49. Leaf, Murray J. "The Punjabi Kinship Terminology as a Semantic System." *AA* 73:3, pp. 545-54.

50. Ibid., 547.

51. Tyler, 1968:358. (Kinship terminology among the Deshastha Brahmans of Maharashtra, mimeo, undated)

52. Gough, 1956:844-9.

53. Raja cites Emeneau, 1953 on this point.

54. Freed, 1963:101.

55. Ibid., 101-2.

56. Vatuk, 1969a:255.

57. E.g. ham baher/ bahen kahtī hăi; măi unko apnī bahen samajhtī hū̃

58. Ibid., 258.

59. Ibid., 261.

60. Ibid.

61. Ibid., 264.

62. Ibid., 266.

63. Ibid., 268.

64. Southworth, 1974, cites evidence from Tamil and Malayalam where structured sets of lexical items are used to mask social distinctions in the society. Full discussion follows in section 7.2.2.4.

65. Cf. Karl Bühler, 1934. *Sprachtheorie*. Jena: Gustav Fischer.

66. Isačenko, A. V., 1964. "On the Conative Function of Language," reprinted in *A Prague School Reader in Linguistics*, Joseph Vachek, ed. Bloomington: Indiana University Press. For a description of these three functions of language in Prague School linguistics see Philip W. Davis, *Modern Theories of Language*, 1973. Englewood Cliffs, N. J., Prentice-Hall, pp. 217-19.

67. See also Austin's essay "Performative Utterances," in *Philosophical Papers* (edited by J. O. Urmson and G. J. Warnock), London: Oxford University Press, 1961, pp. 220-39.

68. Chandrasekhar lists no less than thirteen masculine singular forms (*avan(ōn)*, *ayāḷ*, *addhēham*, *aṅṅorə*, *(aṅṅērə)*, *puḷḷi (-k-karan)*, *mūppar*, *svāmi*, *sār*, *aviṭuńńə*, *tampurān*, *tirumēni*, *tirumanassə*, and *atu)* which can be used "depending on the social status of the referee as well as the social status and the mental attitude of the speaker." (p. 250)

69. A boy of fifteen (B) lives with his father (F), mother (M), elder brother (Eb), and younger brother (Yb). There is a middle-aged servant (S) in the house. They have a friendly neighbor who has a son and his son is a friend (Fr) of the boy. (p. 20)

70. A current treatment of much the same material in Sri Lankan Tamil can be found in Thananjayarajasingham, 1974.

71. *obəvahanse* and *tamunnaanse* are used by laity in addressing monks. The authors do not, however, cite forms used in the reverse situation.

72. *aayusmatun* is used by a senior addressing a junior or by two juniors addressing each other, *stəvirəye* by two high level ordained equal status monks, *haamdururuvənee* by a junior monk addressing a senior, and *unnaanse* is used in addressing a novice or pupil. (p. 91)

73. Thus, for example, *eyaa* is generally used only by and for females with individuals over about 50. Among individuals under 50 it has no such restrictions, and is the most polite pronominal form of address. (p. 92)

74. e.g. *ehey* is generally used only by the laity when talking with monks and by monks when talking with superiors. (p. 94)

75. Few members of the Telugu community are proficient in Koya.

76. Ibid., 255.

77. Ibid.

78. "In these neighborhoods names are generally used in only two contexts. First, given names are used for juniors of approximately equal social status, and for juniors and seniors of much lower social status, in caste or occupational terms. . .Second, given names are used in the case of prominent men of the neighborhood, who are generally referred to by their full name, that is, given name plus 'surname' (e.g., Rām Nivās Vakīl)." (p. 256)

79. This is well illustrated by the following quotations obtained by Vatuk from a 25-year-old Brahman woman:

> There is a woman who lives in the corner house in that alley over there. When we first met I was calling her *bahenjī* because of her age [slightly older than the informant.] But she said to me, "No! You are my nanad [HZ]." I said, "Why?" She said, "Because your *jethani* [HeBW] is a daughter of village Ravi and I am married into Ravi. So I am her *bhābhī* [BW], and therefore I am your *bhābhī* too." So since then I have called her *bhābhī* (HBWBW].
>
> It is the same with the professor's wife next door and the woman across the street. The professor's wife's mother comes from my *sasurāl*, so she is my *nanad* [HZ, i.e., HFZD], and I am her *bhābhī* [BW, i.e. MBSW]. Her children call me *māmī* [MBW]. The woman across the street is married into the village of one of my *jeṭhānīs*, so I consider her my [bhābhī].

80. The following quotation is reported by Vatuk:

> We don't think it is proper to call people by name, so in the *mohallā* we call everyone by some kin term. We can use terms appropriate to the *sasurāl* [husband's village] or to the *pīhar* [woman's natal village]. All of us are newcomers here, so if we choose we can establish *pīhar* kinship between us. For example, our tenants are Baniās. When they moved here she said to me, "Let us make *pīhar* kinship, not *sasurāl* kinship. So we call each other *bahenjī* (eZ) and call each others' husbands *jījājī* (ZH). Our children call them *mausī* and *mausā* and their children call us the same. (p. 266)

81. Jain, 1969:79.

82. Ibid.

83. Ibid.

84. Jain lists the following combinations as being in decreasing order of respect: hakeem doctor + sāhab; hakeem + *jī*; Shri + full name; Mr. + last name (LN); Mr. + first name (FN); bābū + FN; LN + sāhab; LN + bābū; LN + *jī*; FN + *jī*. (p. 81)

85. Thus one can refer to the addressed's home as his *daulatxānā* 'literally, wealth house, palace' and one's own as *γarībxānā* 'humble-house, hut', instead of the common *ghar* 'home, house.' Jain cites twenty-one sets of forms, including verbal, nominal, and adjectival entries, where particular items are restricted to either respectful or non-respectful modes of address.

86. Thus English loan *Doctor* and the Perso-Arabic term *hakeem* accord higher degrees of respect than does H. *vaidya*.

87. Thus *āp kā bhāī dillī gayā hai.* 'Your brother [non-respectful] has gone to Delhi,' ~ *āp ke bhāī dillī gaye hãī*, 'Your brother [respectful] has gone to Delhi.' The latter sentence uses the plural constructions *āp ke bhāī* 'your brothers' and 3rd pl. pres. perfective of √jā 'to go.' The second of these sentences is ambiguous between a plural sense and a singular honorific one. [examples ours]

88. Hindi has three distinct 2nd person pronouns, *tū*, *tum*, and *ap*. The latter two are grammatical plural, and require plural concord features. The choice between *tum* and *āp* either in referring to groups of more than one person, or to an individual whom the speaker wishes to consider 'singular' is determined by a number of sociological factors, including relative age of members of the dyad, reciprocity of their social relationship, presence or absence of outside observation of speech event, degree of formality of situation, etc. For full discussion see Jain 1969, 88-93 and Jain, 1973.

89. As in referring to one's husband by a formula such as *lallū ke pitā jī* 'Lallu's father'. Discussion of this and similar phenomena follows in section 2.3.

90. "In symmetrical intimate relationships most friends (both male or both female) exchange *tū*. In public situations, i.e. in the presence of someone who has a higher social status than that of one of the friends, they use *tum* for each other. Two student friends would use *tum* for each other in the presence of a teacher, so would two clerk-friends in the presence of their boss." (p. 89)

91. For example, Hindi speakers among friends themselves normally address each other with *tū* or *tum*. Jain observes that the use of *āp* within friendship groups is normally taunting or insulting. The insult lies in the breaking of the normal pattern of expectation with regard to pronominal usage. Similarly, a wife who normally is addressed by her husband with *tum*, might be addressed in anger by *tū*. The wife on the other hand, is prohibited through respect conventions of addressing her husband with *tū*, and can only resort to crying. (pp. 89-90)

92. The sentence **ye merā daulatxānā hai* is thus not normally expressible in Hindi because it violates a social prohibition against using the exultory lexical items *daulatxānā* as the head noun in a possessive noun phrase containing the adjectival *merā*. A related constraint renders *ye āp kā γarībxānā hai*, 'this is your hut' inappropriate. (p. 87)

93. e.g. *cey* 'do' (simple command, disrespect); *cey-y-um* 'do' (medial respect); *cey-y-uṅkaḷ* (as singular, respect). (pp. 239-40)

94. e.g. *nii connaay* 'you told' (no formal respect); *niir conniir* 'you told' (medial respect); *conniiṅkaḷ/conniyaḷ* 'you sg. honorific said [or 'you plural said']. (pp. 240-41)

95. Cf. 7.2.2.1.3.

96. e.g. *paalkaaran* 'milkman' (disrespectful) ~ *paalkaaraḷ* (neutral respect). (p. 242)

97. Thus, according to Suseendirarajah, caste names are usually not used as terms of address, but some may be limited for use by certain castes or for according particular levels of respect. e.g. the normal word in Tamil for 'barber' is *ampaṭṭan;* when used by Veeḷaaḷaas, Kooviyass, and some other caste groups, the term of address for a barber is *pariyaaryiyaar*. This last term is more stigmatized than the neutral loan item *barber*. (p. 242)

98. Thus disrespectful *Cuppan* 'a proper name', *aṇṇan* 'elder brother', *ciraapan* 'sheriff'; medial respect *Cuppu, aṇṇai, ciraappu;* respectful *Cuppar, aṇṇar,* and *ciraappar*. (p. 243)

99. Suseendirarajah, 1970:244.

100. Cf. 7.2.2.3.

101. *jab mere father kā death huā. . .* 'When my father's death occured. . .' ~ *jab mere bāp kī maut huī.*

102. Notice in f. 101 the substitution of English *death* for Hindi *maut*.

103. Bharati, 1963:120.

104. Ibid.,

105. Ibid., 127.

106. Jain cites the following anecdote by way of illustration. "When telephoning him (H) in his office I ask for G. L. Aggarwal. Sometimes, the person who receives the call asks, 'Which Aggarwal?'--there being several of them in the office. Then I have to specify and say the (full) name."

107. An extreme example of this is found in the following story cited by Jain (p. 142) which takes place in a hypothetical office:

> "His name is like the one which shines in the sky."
> "Sun?"
> "No, it shines at night."
> "*cā̃dramā*, 'moon'?"
> "Yes, but not exactly."
> "*cā̃d* 'moon'?"
> "Yes, yes."
> "Which one?"
> "The one you said just now."
> "What is his second name?"
> "Like that of god."
> "Which god?" etc.

108. Shanmugam Pillai, 1972:426.

109. Ibid., 430.

110. Southworth, 1974a:178-80.

111. Ibid., 179.

112. Ibid., 182-3.

113. Cf. Jain, 1973 for instances in Hindi and Shanmugam Pillai, 1971 for instances in Tamil. This information is cited by Southworth, 1974:184.

114. Southworth, 1974a:186.

115. An example here will illustrate what we are referring to. There are a number of expressions in Hindi which are usually thought to indicate positive responses of some type or another, and which are generally translated as 'yes' (*jī hā̃, hā̃ jī, hā̃ jī*). These expressions are usually accompanied by a simultaneous nodding or rolling of the head. Yet treating these utterances as purely referential in nature misses much of their semantics, and opens the way for cultural misinterpretation. These constructions are generally used to indicate that the listener is following the flow of the speaker's narrative, and is still an active participant in the

conversation. These expressions, thus frequently maintain the continuity of a speech event, but do not necessarily indicate agreement with the propositional content of what the speaker is saying.

ABBREVIATIONS USED IN BIBLIOGRAPHY

AA	*American Anthropologist.*
AL	*Anthropological Linguistics.*
AO	*Archiv Orientální.*
CA	*Cognitive Anthropology.* Edited by Stephen A. Tyler. New York: Holt, Rinehart and Winston, 1969.
CCSAL	*Contact and Convergence in South Asian Languages.* Edited by Franklin C. Southworth and Mahadev L. Apte. *International Journal of Dravidian Linguistics* 3:1 (1974).
CTL5	*Current Trends in Linguistics,* volume 5, *Linguistics in South Asia.* Edited by Thomas A. Sebeok. The Hague: Mouton & Co., 1969.
DLEF	Murray B. Emeneau. *Dravidian Linguistics, Ethnology, and Folktales: Collected Papers.* Annamalainagar: Annamalai University Linguistics Department Publication No. 8, 1967.
IJAL	*International Journal of American Linguistics.*
IJDL	*International Journal of Dravidian Linguistics.*
IL	*Indian Linguistics.*
Lg	*Language.*
LDSA	*Linguistic Diversity in South Asia.* Edited by Charles A. Ferguson and John J. Gumperz. *International Journal of American Linguistics* 26:3 (1960) pt. 3.
LSI	George Abraham Grierson. *Linguistic Survey of India.* 11 volumes. Reprinted edition, Delhi: Motilal Banarsidass, 1967.
LSIT	*Language and Society in India: Transactions.* Simla: Indian Institute of Advanced Study, volume 8, 1969.
LSG	John J. Gumperz. *Language in Social Groups.* Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1971.
NWAVE	*New Ways of Analyzing Variation in English.* Edited by Charles-James N. Bailey and Roger W. Shuy. Washington, D. C.: Georgetown University Press, 1973.
PAPS	*Proceedings of the American Philosophical Society.*
PFICSTS	*Proceedings of the First International Conference-Seminar of Tamil Studies (1966).* Edited by Father Xavier S. Thani Nayagam. Kuala Lumpur: International Association of Tamil Research, 1969.
RSL	*Readings in the Sociology of Language.* Edited by Joshua A. Fishman. The Hague: Mouton & Co., 1968.

SCIS	*Structure and Change in Indian Society*. Edited by Milton Singer and Bernard S. Cohen. Chicago: Aldine Publishing Company, 1968.
SIL	*Studies in Indian Linguistics*. Edited by Bh. Krishnamurti. Poona and Annamalainagar: Centres of Advanced Study in Linguistics, 1968.
TSDL	*Third Seminar on Dravidian Linguistics (Annamalai University, 1971)*. Edited by S. Agesthialingom and S. V. Shanmugam. Annamalainagar: Annamalai University, Department of Linguistics, Publication No. 27, 1972.

BIBLIOGRAPHY

Ahmad, S. Maqbul. 1969. "The Position of Urdu in India Today." In *LSIT*, pp. 147-151.

Ananthanarayana, H. S. 1968. "Sound Changes in Sanketi, A Tamil Dialect." In *SIL*, pp. 1-14.

Andronov, M. S. 1962. *Razgovornyj Tamil'skij Jazyk i ego Dialekty* (Colloquial Tamil and its Dialects). Moscow: Izdatel'stvo Vostočnoj Literatury.

__________. 1963. "Dravidian Languages." *AO* 31:2.177-97.

__________. 1964a. "Lexicostatistic Analysis of the Chronology of Disintegration of Proto-Dravidian." *Indo-Iranian Journal* 7:2/3.170-86.

__________. 1964b. "On the Typological Similarity of New Indo-Aryan and Dravidian." *IL* 25.119-26. Reprinted as "Dravidian and Aryan: From the Typological Similarity to the Similarity of Forms." In *Two Lectures on the Historicity of Language Families*, pp. 1-13. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 15 (1968).

__________. 1970. *Dravidian Languages*. Translated by D. M. Segal. Moscow: "Nauka" Publishing House. (Original in Russian, 1965).

Apte, M. L. 1962. "Linguistic Acculturation and its Relation to Urbanization and Socio-Economic Factors." *IL* 23.5-25.

__________. 1974. "Pidginization of a Lingua Franca: A Linguistic Analysis of Hindi-Urdu Spoken in Bombay." In *CCSAL*, pp. 21-41.

Arasaratnam, S. 1967. "Nationalism, Communalism and National Unity in Ceylon." In *India and Ceylon: Unity and Diversity*, pp. 260-78. Edited by Phillip Mason. London: Oxford University Press.

Asher, R. E. 1972. "Dravidian Separateness: Invention or Reality." *South Asian Review*, 6:1.33-42.

Aziz, Ahmad. 1964. *Studies in Islamic Culture in the Indian Environment*. Oxford: Clarendon Press.

Bailey, Charles-James N. 1972. "The Patterning of Language Variation." In *Varieties of Present-Day English*. Edited by R. W. Bailey and J. L. Robinson. New York: Macmillan Co.

__________. 1973a. *Variation and Linguistic Theory*. Washington, D. C.: Center for Applied Linguistics.

__________. 1973b. "Variation Resulting from Different Rule Orderings in English Phonology." In *NWAVE*, pp. 211-52.

Bansal, R. K. 1969. *The Intelligibility of Indian English*. Hyderabad: Da'iratu'l-Ma'arif Press.

Bazaz, Prem Nath. 1969. "The Problem of Languages in India." In *LSIT*, pp. 152-8.

Beames, John. 1960. *Outlines of Indian Philology and Other Philological Papers.* Calcutta: Bholanath Hazra at Rupabani Press. Reprinted from *Indian Studies: Past and Present* 1:2 (1960) 333-76. (Original 1866-7).

Bean, Susan S. 1974. "Linguistic Variation and the Caste System in South Asia." *IL* 35:4.277-93.

Becker, Henrik. 1948. *Der Sprachbund.* Leipzig: Humboldt-Bücherei.

Bendix, Edward H. 1974. "Indo-Aryan and Tibeto-Burman Contact as Seen Through Nepali and Newari Verb Tenses." *CCSAL*, pp. 42-59.

Benedict, Paul K. 1972. *Sino-Tibetan: A Conspectus.* London: Cambridge University Press.

Benveniste, Émile. 1969. *Le Vocabulaire des Institutions Indo-Européens.* Paris: Les Editions de Minuit.

Bernstein, Basil. 1964. "Elaborated and Restricted Codes: Their Social Origins and Some Consequences." In *The Ethnography of Communication*, pp. 55-69. Edited by John J. Gumperz and Dell Hymes. *AA* 66:6 pt. 2.

Bharati, A. 1963. "Kinship Term Avoidance and Substitution in North Indian Middle Class Milieux." *Sociologus* New Series 13:2.112-20.

Bhattacharya, S. 1957. *Ollari: A Dravidian Speech.* (Department of Anthropology, Government of India, Memoir No. 3, 1956). Delhi: Manager of Publications.

__________. 1966. "Some Munda Etymologies." In *Studies in Comparative Austroasiatic Linguistics*, pp. 28-40. Edited by Norman Zide. The Hague: Mouton & Co.

__________. 1970. "Kinship Terms in the Munda Languages." *Anthropos* 65:3/4.444-65.

__________. 1972. "Dravidian and Munda (A Good Field for Areal and Typological Studies)." In *TSDL*, pp. 241-56.

Bloch, Jules. 1910. "Castes et Dialectes en Tamoul." *Memoires de la Société de Linguistique de Paris* 16.1-30.

__________. 1914. *The Formation of the Marāthī Language.* Translated by Dr. Dev Raj Chanana. Delhi: Motilal Banarsidass, pp. 1-38, 287-9. [Citations from 1970 edition].

__________. 1924. "Sanskrit et Dravidien." *Bulletin de la Société de Linguistique de Paris* 25:1.1-21. Reprinted in *Pre-Aryan and Pre-Dravidian in India*, pp. 35-59. Edited by Sylvain Levi, Jean Przyluski and Jules Bloch. Translated by Prabodh Chandra Bagchi. Calcutta: University of Calcutta, 1929.

__________. 1934. *Indo-Aryan: From the Vedas to Modern Times.* Translated by Alfred Master. Paris: Librairie Adrien-Maisonneuve, 1965, pp. 321-31. (Original in French).

Bloomfield, Leonard. 1926. "A Set of Postulates for the Science of Language." *Lg* 2.153-64. Reprinted in *Readings in Linguistics I.* Edited by Martin Joos. Fourth edition. Chicago: University of Chicago Press (1966), pp. 26-31. [Citations from Joos, 1966.]

__________. 1933. *Language.* New York: Holt, Rinehart and Winston.

Boas, Franz. 1920. "The Classification of American Languages." *AA* 22:4.367-76. Reprinted in *Race, Language and Culture*, pp. 211-18. Franz Boas. New York: MacMillan and Company, 1940.

__________. 1929. "Classification of American Indian Languages." *Lg* 5:1.1-7. Reprinted in *Race, Language and Culture*, pp. 219-25. Franz Boas. New York: MacMillan and Company, 1940.

Bonfante, G., and Sebeok, Thomas A. 1944. "Linguistics and the Age and Area Hypothesis." *AA* 46:3.382-6.

Bottiglioni, Gino. 1954. "Linguistic Geography: Achievements, Methods, and Orientations." *Word* 10:2/3.75-87.

Brandreth, E. L. 1878. "On the Non-Aryan Languages of India." *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 10.1-32.

Brass, Paul R. 1974. *Language, Religion and Politics in North India.* New York: Cambridge University Press.

Bray, Denys De S. 1909. *The Brahui Language,* Part 1, *Introduction and Grammar.* Calcutta: Superintendent, Government Printing.

Bright, William. 1960a. "Social Dialect and Language History." *Current Anthropology* 1:5/6.424-5.

__________. 1960b. "A Study of Caste and Dialect in Mysore." *IL* 21. 45-50. Revised and printed as "Linguistic Change in Some Indian Caste Dialects." In *LDSA*, pp. 19-26.

__________. 1966a. "Dravidian Metaphony." *Lg* 42:2.311-22.

__________. 1966b. "Language, Social Stratification, and Cognitive Orientation." *Sociological Inquiry* 36:2.313-18.

__________. 1968. "Social Dialect and Semantic Structure in South Asia." In *SCIS*, pp. 455-60.

__________, and A. K. Ramanujan. 1962. "Sociolinguistic Variation and Language Change." In *Proceedings of the Ninth International Congress of Linguists (1962)*, pp. 1107-13. Edited by Horace G. Lunt. The Hague: Mouton & Co.

Brown, Roger. 1958. *Words and Things.* New York: The Free Press.

__________, and Gilman, Albert. 1960. "The Pronouns of Power and Solidarity." In *Style in Language*, pp. 253-76. Edited by Thomas A. Sebeok. Cambridge, Mass.: The Technology Press of M.I.T.; New York: John Wiley & Sons, Inc. Reprinted in *RSL*, pp. 252-75.

Brown, W. Norman. 1952. "The Language Problem of India and Pakistan." *Report of the Third Annual Round Table Meeting on Linguistics and Language Teaching*, pp. 17-30. Edited by Salvatore J. Castiglione. Georgetown University Monograph Series No. 25, Washington, D. C., Georgetown University Press.

Burling, Robbins. 1967. *Proto-Lolo-Burmese*. Bloomington: Indiana University Publications, vol. 43.

Burrow, T. 1970. *The Pengo Language (Grammar, Texts, and Vocabulary)*. Oxford: Clarendon Press.

__________, and Bhattacharya, S. 1953. *The Parji Language: A Dravidian Language of Bastar*. Hertford: Stephen Austin and Sons, Ltd.

__________, and Emeneau, M. B. 1961. *A Dravidian Etymological Dictionary*. London: Oxford University Press.

__________. 1968. *A Dravidian Etymological Dictionary: Supplement*. London: Oxford University Press.

Caldwell, Rt. Rev. Robert. 1856. *A Comparative Grammar of the Dravidian or South-Indian Family of Languages*, 3rd ed. Revised and edited by Rev. J. L. Wyatt and T. Ramakrishna Pillai. Reprint ed., Madras: University of Madras, 1961.

Cardona, George. 1974. "The Indo-Aryan Languages." *Encyclopaedia Britannica*. 15th ed. Vol. 9, 439-50.

Cassidy, Frederic Gomes. 1961. *Jamaican Talk: Three Hundred Years of the English Language in Jamaica*. London: MacMillan & Company, Ltd.

Cedergren, Henrietta J. 1973. "On the Nature of Variable Constraints." In *NWAVE*, pp. 13-22.

Chagla, M. C. 1967. *Chagla on Language and Unity*. Bombay: Popular Prakashan.

Chandrasekhar, A. 1970. "Personal Pronouns and Pronominal Forms in Malayalam." *AL* 12:7.246-55.

Chatterji, S. K. 1926. *The Origin and Development of the Bengali Language*. Part I, *Introduction, Phonology*. Reprint ed., London, George Allen and Unwin, Ltd., 1970, pp. 1-178, 189-223. [Citations from 1970 edition].

__________. 1931. "Calcutta Hindustani: A Study of a Jargon Dialect." *IL* 1:2-4. Reprinted in *Select Papers*, pp. 204-56. New Delhi: People's Publishing House, 1972.

__________. 1942. *Indo-Aryan and Hindi.* Ahmedabad: Gujarat Vernacular Society; reprint ed., Calcutta: Firma K. L. Mukhopadhyay, 1960 and 1969.

__________. 1953. "The Indian Synthesis, and Racial and Cultural Inter-Mixture in India." (Presidential Address) *Proceedings and Transactions of the All India Oriental Conference, Seventeenth Session, Ahmedabad* Section III, pp. 22-60.

Chernyshev, V. A. 1971. "Nekotorye cherty Bombeiskogo Govora Khindustani, (na materiale sovremennoi prozy khindi)." [Some features of Bombay Hindustani speech (in modern Hindi prose material)]. In *Indiiskaia i Iranskaia Filologiia (Voprosy Dialektologii)*, pp. 121-41. Edited by N. A. Dvoriankov. Moscow: Izdatel'stvo "Nauka".

Chickermane, D. V. 1971. *Impact of Bilingualism on the Progress of Children in Primary Schools in Rural Areas.* New Delhi: National Council of Educational Research and Training.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax.* Cambridge, Mass.: M.I.T. Press.

Cingaram, K. 1968. "The Consonantal System of Tigaḷa, a Tamil Dialect." In *SIL*, pp. 75-80.

Coates, William Ames. 1961. "The Languages of Ceylon in 1946 and 1953." *University of Ceylon Review* 19.81-91.

Cohn, B. S. 1957. "India as a Racial, Linguistic, and Cultural Area." In *Introducing India in Liberal Education*, pp. 51-68. Edited by Milton Singer. Chicago: University of Chicago Press.

Commissioner for Linguistic Minorities in India, Report of the Commissioner. 1972. *The Twelfth Report of the Commissioner. (July, 1969-June, 1970).* New Delhi: Government of India Press.

Committee on Broadcasting and Information Media, Report of the Committee. 1966. *Broadcasting and Information Media.* New Delhi: Government of India Press.

Committee of Members of Parliament on Education, Report of the Committee. 1967. *Report on Education.* Faridabad: Government of India Press.

Dalgado, Sebastião Rodolfo. 1900. *Dialecto Indo-Português de Ceylâõ.* Lisboa: Imprensa Nacional.

__________. 1906. "Dialecto Indo-Português do Norte." *Revista Lusitana* 9.142-66.

__________. 1917. "Dialecto Indo-Português de Negapatão." *Revista Lusitana* 20.40-53.

Das, Sisir Kumar. 1968. "Forms of Address and Terms of Reference in Bengali." *AL* 10:4.19-31.

Das Gupta, J. 1969. "Official Language Problems and Policies in South Asia." In *CTL5*, pp. 578-96.

__________. 1970. *Language Conflict and National Development.* Berkeley: University of California Press.

__________. 1975. "Religion, Language and Social Mobilization." In *Can Language be Planned?*, pp. 53-64. Edited by Joan Rubin and Björn H. Jermudd. Honolulu: The University Press of Hawaii.

__________ and Gumperz, John J. 1968. "Language, Communication and Control in North India." In *Language Problems of Developing Nations*, pp. 17-26. Edited by Joshua A. Fishman, Charles A. Ferguson, and Jyotirindra Das Gupta. New York: John Wiley and Sons.

Davidson, T. T. L. 1969. "Indian Bilingualism and the Evidence of the Census of 1961." *Lingua* 22.176-96.

Davis, Kingsley and Warner, W. Lloyd. 1937. "Structural Analysis of Kinship." *AA* 39:2.291-313.

DeCamp, David. 1961. "Social and Geographical Factors in Jamaican Dialects." In *Proceedings of the Conference on Creole Language Studies* (1959), (Creole Language Studies, No. 2), pp. 61-84. Edited by R. B. LePage. London: Macmillan & Co., Ltd.

__________. 1971. "Toward a Generative Analysis of a Post-Creole Speech Continuum." In *Pidginization and Creolization of Languages (Proceedings of a Conference, Jamaica, 1968)*, pp. 349-70. Edited by Dell Hymes. London: Cambridge University Press.

De Silva, M. W. S. 1974. "Convergence in Diglossia: The Sinhalese Situation." In *CCSAL*, pp. 60-91.

Deshmukh, C. D. 1972. *Hindustani: Rashtra bhasha or Lingua Franca?* Bombay: Mahatma Gandhi Memorial Research Centre.

Dharmadasa, K. N. O. 1974. "The Creolization of an Aboriginal Language: The Case of Vedda in Sri Lanka (Ceylon)." *AL* 16:2.79-106.

Diebold, A. Richard. 1961. "Incipient Bilingualism." *Lg* 37.97-112.

Diffloth, Gerard F. 1968. "The Irula Language, A Close Relative of Tamil." Unpublished Ph.D. dissertation, U. C. L. A.

Dimock, Edward C. 1957. "Symbolic Forms in Bengali." *Bulletin of the Deccan College Research Institute* (Taraporewala Memorial Volume) 18:22-9.

__________. 1960. "Literary and Colloquial Bengali in Modern Bengali Prose." In *LDSA*, pp. 43-63.

Dumont, Louis. 1953. "The Dravidian Kinship Terminology as an Expression of Marriage." *Man* 53.34-9. Reviewed by A. R. Radcliffe-Brown in *Man* 53 (1953), 112. Reply by Dumont in *Man* 53 (1953), 143.

Education Commission, Report of the Commission. 1967a. *Comments and Reactions*. New Delhi: Lok Sabha Secretariat.

__________. 1967b. *Summary of Recommendations*. New Delhi: National Council of Educational Research and Training.

Elizarenkova, T. 1972. "Influence of Dravidian Phonological System on Sinhalese." *IJDL* 1:2.126-37.

__________. 1938a. "Echo-Words in Toda." *New Indian Antiquary* 1.109-17. Reprinted in *DLEF*, pp. 37-45.

__________. 1938b. "Kinship and Marriage Among the Coorgs." *Journal of the Royal Asiatic Society of Bengal* 4:1.123-47. Reprinted in *DLEF*, pp. 333-56.

__________. 1941. "Language and Social Forms: A Study of Toda Kinship Terms and Dual Descent." In *Language, Culture, and Personality: Essays in Memory of Edward Sapir*, pp. 158-79. Edited by Leslie Spier. Menasha, Wisc.: Sapir Memorial Publishing Fund. Reprinted in *DLEF*, pp. 233-57.

__________. 1944. *Kota Texts*. University of California Publications in Linguistics, vol. 2. Berkeley: University of California Press.

__________. 1948a. "More on the Dravidian Etymologies in 'Taboos on Animal Names'." *Lg* 24.180-1. Reprinted in *DLEF*, p. 210.

__________. 1948b. "Taboos on Animal Names." *Lg* 24.56-63. Reprinted in *DLEF*, pp. 201-9.

__________. 1949. "Language and Non-linguistic Patterns." Presidential address, read at the 24th annual meeting of the Linguistic Society of America, Philadelphia, 28 December. Reprinted in *DLEF*, pp. 211-23.

__________. 1953. "Dravidian Kinship Terms." *Lg* 29.339-53. Reprinted in *DLEF*, pp. 123-38.

__________. 1954. "Linguistic Pre-History of India." *Proceedings of the American Philosophical Society* 98:4.282-92.

__________. 1955. *Kolami: A Dravidian Language*. University of California Publications in Linguistics, vol. 12. Berkeley: University of California Press. Reprint ed., Annamalainagar: Annamalai University Department of Linguistics Publication No. 2, 1961.

__________. 1956. "India as a Linguistic Area." *Lg*.32:1.3-16. Reprinted in *Language in Culture and Society*, 642-53. Dell Hymes. New York: Harper and Row, 1964. Also in *DLEF*, pp. 172-86.

__________. 1962a. "Bilingualism and Structural Borrowing." *PAPS* 106:5.430-42.

__________. 1962b. *Brahui and Dravidian Comparative Grammar*. University of California Publications in Linguistics, vol. 27. Berkeley: University of California Press.

__________. 1967. "The South Dravidian Languages." *Journal of the American Oriental Society* 87:4.365-413. Summarized in *Proceedings of the First International Conference Seminar on Tamil Studies (1966)*, vol. 2, pp. 563-71. Kuala Lumpur: International Association of Tamil Research, 1969.

__________. 1969. "Onomatopoetics in the Indian Linguistic Area." *Lg* 45:2, pt. 1.274-99.

__________. 1971. "Dravidian and Indo-Aryan: The Indian Linguistic Area." *Symposium on Dravidian Civilization (1968)*. Edited by Andrée Sjoberg. Austin: Center for Asian Studies of the University of Texas at Austin, Asian Series No. 1.

__________. 1974a. "The Indian Linguistic Area Revisited." *CCSAL*, pp. 92-134.

__________. 1974b. "Ritual Structure and Language Structure of the Todas." *Transactions of the American Philosophical Society*, New Series 64:6.

Emmerick, R. E. 1974. "Iranian Languages." *Encyclopedia Britannica* 15th ed., vol. 9, 450-7.

English Review Committee, Report of the Committee. 1965. *Report*. New Delhi: University Grants Commission.

Ferguson, Charles A. 1959. "Diglossia." *Word* 15:2.325-40. Reprinted in *Language in Culture and Society*, pp. 429-37. Edited by Dell Hymes. New York: Harper and Row, 1964. Also in *Language Structure and Language Use: Essays by Charles A. Ferguson*, pp. 1-26. Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1971.

__________, and Gumperz, John J., eds. 1960. *Linguistic Diversity in South Asia: Studies in Regional, Social and Functional Variation. IJAL* 26:3 pt. 2.

Fischer, John L. 1958. "Social Influences on the Choice of a Linguistic Variant." *Word* 14:1.47-56. Reprinted in *Language in Culture and Society*, pp. 483-8. Edited by Dell Hymes. New York: Harper and Row, 1964.

Fishman, Joshua A. 1967. "Bilingualism with and without Diglossia; Diglossia with and without Bilingualism." *Journal of Social Issues* 23:2. 29-37. Revised and printed as "Societal Bilingualism: Stable and Transitional." In *Sociolinguistics: A Brief Introduction*, pp. 73-89. By Joshua A. Fishman. Rowley, Mass.: Newbury House Publishers, 1970. Revised and Printed in *Language in Sociocultural Change: Essays by Joshua A. Fishman*, pp. 135-52. Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1972.

__________. 1972. *Language in Socio-Cultural Change*. Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press. Chapter 5, "Societal Bilingualism."

Fonseca, Fernando V. Peixoto Da. 1959. *Noções de Historia da Lingua Portuguesa*. Lisbon: Livraria Classica Editora.

Forbes, C. J. F. S. 1878. "On Tibeto-Burman Languages." *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 10.210-27.

Frake, Charles O. 1962. "The Ethnographic Study of Cognitive Systems." In *Anthropology and Human Behavior*, pp. 72-85. Edited by Thomas Gladwin and William C. Sturtevant. Washington, D. C.: Anthropological Society of Washington. Reprinted in *RSL*, pp. 434-46. Also in *CA*, pp. 28-41.

Freed, Stanley A. 1963. "Fictive Kinship in a North Indian Village." *Ethnology* 2:1.86-103.

Freeman, David M. 1968. "Value Aggregation and the Ceylonese Language Dispute." Unpublished Ph.D. dissertation, University of Denver.

Friedrich, Paul W. 1961. "Review of *Linguistic Diversity in South Asia*." *Lg* 37:1.163-8.

__________. 1962. "Language and Politics in India." *Daedalus* 91:3. 563-9. Reprinted in University of Chicago Committee of South Asian Studies No. 24, 1962.

Gair, James W. 1968. "Sinhalese Diglossia." *AL* 10:8.1-15.

Geiger, Wilhelm and Kuhn, Ernst, eds. 1901. *Grundriss der Iranischen Philologie.* (2 vols.) Strassburg: Verlag Von Karl J. Trübner.

Ghurye, G. S. 1968. *Social Tensions in India.* Bombay: Popular Prakashan, pp. 430-84.

Gopal, Ram. 1966. *Linguistic Affairs of India.* Bombay: Asia Publishing House.

Gough, Kathleen E. 1956. "Brahman Kinship in a Tamil Village." *AA* 58:5.826-53.

Greenberg, Joseph H. 1956. "The Measurement of Linguistic Diversity." *Lg* 32:1.109-15. Reprinted in *Language, Culture, and Communication: Essays by Joseph H. Greenberg*, pp. 68-77. Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1971.

__________. 1966. "Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements." In *Universals of Language*, pp. 73-113. Edited by Joseph H. Greenberg. Cambridge, Mass.: M.I.T. Press.

Grierson, George Abraham. 1903-28. *Linguistic Survey of India.* 11 vols. Reprint ed., Delhi: Motilal Banarsidass, 1967.

__________. 1906. *The Piśāca Languages of North-Western India.* London: Royal Asiatic Society (Monograph No. 8). Reprint ed., New Delhi: Munshiram Manoharlal Oriental Publishers, 1969.

__________. 1915a. "The Linguistic Classification of Kashmiri." *Indian Antiquary: A Journal of Oriental Research* 44.257-70.

__________. 1915b. "The North-Western Group of the Indo-Aryan Vernaculars." *Indian Antiquary: A Journal of Oriental Research* 44.226-8.

Gumperz, John J. 1958. "Dialect Differences and Social Stratification in a North Indian Village." *AA* 60:4.668-82. Reprinted in *LSG*, pp. 25-47.

__________. 1961. "Speech Variation and the Study of Indian Civilization." *AA* 63:5 pt. 1.976-88.

__________. 1964a. "Hindi-Punjabi Code Switching in Delhi." In *Proceedings of the Ninth International Congress of Linguists*, pp. 1115-24. Edited by H. Lunt. The Hague: Mouton and Co. Reprinted in *LSG*, pp. 205-19.

__________. 1964b. "Linguistic and Social Interaction in Two Communities." In *The Ethnography of Communication*, pp. 137-53. Edited by John J. Gumperz and Dell Hymes. *AA* 66:6 pt. 2. Reprinted in *LSG*, 151-76.

__________. 1967. "Language and Communication." In *Social Goals and Indicators for American Society*, vol. II, pp. 219-31. Edited by Bertram M. Gross. *The Annals of the American Academy of Political and Social Science* 373.

__________. 1968. "Some Desiderata in South Asian Areal Linguistics." *SIL*, pp. 118-23.

__________. 1969a. "How Can We Describe and Measure the Behavior of Bilingual Groups?" In *Description and Measurement of Bilingualism: An International Seminar (1967)*, pp. 242-9. Edited by L. G. Kelly. Toronto: University of Toronto Press.

__________. 1969b. "Sociolinguistics in South Asia." In *CTL5*, pp. 597-606.

__________, and Naim, C. M. 1960. "Formal and Informal Standards in the Hindi Regional Language Area." *LDSA*, pp. 92-118. Reprinted in *LSG*, pp. 48-76.

__________, and Wilson, R. 1971. "Convergence and Creolization: A Case From the Indo-Aryan/Dravidian Border in India." In *Pidginization and Creolization of Languages*, pp. 151-67. Edited by Dell H. Hymes. London: Cambridge University Press. Reprinted in *LSG*, pp. 251-73.

Haggi, S. A. H. 1972. "Language and Union-State Relationships." In *Union and the States*, pp. 367-9. Edited by S. N. Jain and Subhash C. Kashyap, et. al. Delhi: National Publishing House.

Hai, Muhammad Abdul. 1975. "The Development of Bengali Since the Establishment of Pakistan." In *Can Language Be Planned?*, pp. 189-94. Edited by Joan Rubin and Björn H. Jermudd. Honolulu: The University of Hawaii Press.

Hall, Robert A., Jr. 1946. "Bartoli's 'Neolinguistica'." *Lg* 22:4.273-83.

Halverson, John. 1966. "Prolegomena to the Study of Ceylon English." *University of Ceylon Review* 24:1 and 2.61-75.

Hardgrave, R. L. Jr. 1965. *The Dravidian Movement.* Bombay: Popular Prakashan.

Harrison, Sellig S. 1960. *India: The Most Dangerous Decades.* New Jersey: Princeton University Press.

Haugen, Einar. 1952. "The Impact of English on American-Norwegian Letter Writing." In *Studies in Honor of Albert Morey Sturtevant*, pp. 76-102. Edited by Clyde Kenneth Hyder. Lawrence, Kansas: University of Kansas Publications, Humanistic Studies, No. 29.

__________. 1953. *The Norwegian Language in America: A Study in Bilingual Behavior.* 2 vols. Philadelphia: University of Pennsylvania Press.

__________. 1956. *Bilingualism in the Americas: A Bibliography and Research Guide.* University of Alabama: American Dialect Society, Publication No. 26.

__________. 1966a. "Dialect, Language, Nation." *AA* 68:4.922-35. Reprinted in *The Ecology of Language: Essays by Einar Haugen*, pp. 237-53. Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1972.

__________. 1966b. "Semicommunication: the Language Gap in Scandinavia." *Sociolinguistic Inquiry* 36.280-97. Reprinted in *The Ecology of Language: Essays by Einar Haugen, pp. 215-36.* Selected and introduced by Anwar S. Dil. Stanford: Stanford University Press, 1972.

Hesseling, Dirk C. 1899. *Het Afrikaans; Bijdrage tot de Geschiedenis der Nederlandse taal in Zuid-Afrika.* Leiden: E. J. Brill.

Hodson, T. C. 1936. "Bilingualism in India." *Transactions of the Philological Society of London*, pp. 85-91. London: David Nutt (A. G. Berry).

Hoffmann, Karl. 1952. "Wiederholende Onomatopoetika im Altindischen." *Indogermanische Forschungen* 60.254-64.

Holmer, Nils M. 1963. "The Morphological Structure of Austro-asiatic Languages." In *Linguistic Comparison in South East Asia and the Pacific*, pp. 17-23. Edited by H. L. Shorto. London: School of Oriental and African Studies.

Hymes, Dell H. 1962. "The Ethnography of Speaking." In *Anthropology and Human Behavior*, pp. 13-53. Edited by Thomas Gladwin and William C. Sturtevant. Washington, D. C.: Anthropological Society of Washington. Reprinted in *RSL*, pp. 99-138. Also in *Language, Culture and Society*, pp. 189-223. Edited by Ben G. Blount. Cambridge, Mass.: Winthrop Publishers, Inc., 1974.

__________, ed. 1971. *Pidginization and Creolization of Languages (Proceedings of a Conference, Jamaica, 1968)*. London: Cambridge University Press.

Irschick, Eugene F. n.d. "Dravidianism in South Indian Politics." (mimeo)

Ivić, Pavle. 1964. "Structure and Typology of Dialectal Differentiation." In *Proceedings of the Ninth International Congress of Linguists*, pp. 115-28. Edited by Horace G. Lunt. The Hague: Mouton & Co.

Jaberg, Karl. 1936. *Aspects géographiques du langage*. Paris: Librairie E. Droz.

Jakobson, Roman. 1931. "Uber die phonologischen Sprachbünde." *Travaux du Cercle Linguistique de Prague* 4.234-40. Reprinted in *Selected Writings, vol. I, Phonological Studies*, pp. 137-43. Roman Jakobson. The Hague: Mouton & Co., 1962.

__________. 1936. "Sur la théorie des affinités phonologiques entre les langues." *Actes du Quatrième Congrès International de Linguistes*. Copenhagen: Munksgaard, 1938. Reprinted in Trubetzkoy, N. S. *Principes de Phonologie*, pp. 351-65. Translated by J. Cantineau. Paris: Klincksieck, 1949. Also in Jakobson, Roman, *Selected Writings, vol. I, Phonological Studies*, pp. 234-46. The Hague: Mouton & Co., 1962.

__________, Fant, C. Gunnar M., and Halle, Morris. 1951. *Preliminaries to Speech Analysis*. Cambridge, Mass.: M. I. T. Press.

Jain, Dhanesh Kumar. 1969. "Verbalization of Respect in Hindi." *AL* 11:3.79-97.

__________. 1973. "Pronominal Usage in Hindi: A Sociolinguistic Study." Unpublished Ph.D. dissertation; University of Pennsylvania.

Kachru, Braj B. 1965. "The Indianness of Indian English." *Word* 21:3. 391-410.

__________. 1966. "Indian English: A Study in Contextualization." In *In Memory of J. R. Firth*, pp. 255-87. Edited by C. E. Bazell, J. C. Catford, et. al. London: Longmans, Green and Co., Ltd.

__________. 1969. "English in South Asia." In *CTL5*, pp. 627-78.

Katmatchinathan, A. 1969. *The Tirunelvēli Tamil Dialect*. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 19.

Kanapathi Pillai, K. 1958. "The Jaffna Dialect of Tamil: A Phonological Study." *IL* 19 (Turner Jubilee Volume I).219-27.

Karunakaran, K. 1971. *The Kollimalai Tamil Dialect*. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 26.

Karunatilake, W. S. and Suseendirarajah, S. 1973. "Phonology of Sinhalese and Sri Lanka Tamil--A Study in Contrast and Interference." *IL* 34:3.180-90.

__________. 1975. "Pronouns of Address in Tamil and Sinhalese--A Sociolinguistic Study." *IJDL* 4:1.83-96.

Karve, Irawati. 1965. *Kinship Organization in India*. Second Revised Edition, Bombay, New York: Asia Publishing House (original, 1953).

__________. 1970. *Kinship Patterns, Terminology and Behavior in India*. Miami: Field Research Projects, Occasional Paper No. 10.

Katz, Jerrold J. 1966. *The Philosophy of Language*. New York: Harper and Row.

__________. 1972. *Semantic Theory*. New York: Harper and Row.

Katz, Jerrold J. and Postal, Paul M. 1964. *An Integrated Theory of Linguistic Descriptions*. Cambridge, Mass.: M. I. T. Press.

Kearney, Robert N. 1967. *Communalism and Language in the Politics of Ceylon*. Durham: N. C.: Duke University Press.

Kelkar, Ashok R. 1957. "'Marathi English': A Study in Foreign Accent." *Word* 13.268-82.

Kelley, Gerald. 1969. "Telugu." In *CTL5*, pp. 382-93.

Kelly, L. G., ed. 1969. *Description and Measurement of Bilingualism: An International Seminar*. Toronto: University of Toronto Press.

Killingley, D. H. 1969. "Retroflexion in Sanskrit." In *PFICSTS*, pp. 596-606.

Knowlton, E. C. 1969. "Portuguese-Tamil Linguistic Contacts." In *PFICSTS*, pp. 607-17.

Kodesia, Krishna. 1969. *The Problems of Linguistic States In India*. Delhi: Sterling Publishers Pvt. Ltd.

Konow, Sten. 1902. "Note on the Languages Spoken Between the Assam Valley and Tibet." *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 127-37.

__________. 1904. "The Kurku Dialect of Munda Family of Speech." *The Journal of the Royal Asiatic Society of Great Britain and Ireland* 423-33.

__________. 1908. "Notes on the Munda Family of Speech in India." *Anthropos* 3.68-82.

Kothari, Rajni. 1970. *Politics in India*. Delhi: Orient Longman Ltd.

Krishnamurti, Bh. 1961. *Telugu Verbal Bases: A Comparative and Descriptive Study*. University of California Publications in Linguistics, vol. 24. Berkeley: University of California Press.

__________. 1962. *Māṇḍalika vrittipadakōśam*, vol. 1, *Agriculture*. Hyderabad: Andhra Pradesh Sahitya Akademi.

__________. 1969a. "Comparative Dravidian Studies." In *CTL5*, pp. 309-33.

__________. 1969b. *Koṇḍa or Kūbi, A Dravidian Language (Texts, Grammar, and Vocabulary)*. Hyderabad: Tribal Cultural Research and Training Institute, Government of Andhra Pradesh.

Kuiper, F. B. J. 1948. *Proto-Munda Words in Sanskrit*. Amsterdam: Noord-Hollandsche Uitg. Mij.

__________. 1962. "Notes on Old Tamil and Jaffna Tamil." *Indo-Iranian Journal* 5:1.52-64.

__________. 1967. "The Genesis of a Linguistic Area." *Indo-Iranian Journal* 10:2/3.81-102. Reprinted in *CCSAL*, pp. 135-53. [citations from *CCSAL*]

Kumaraswami Raja, N. 1972. "Personal Kinship Terms." In *TSDL*, pp. 123-34.

Kurath, Hans. 1939. *Handbook of the Linguistic Geography of New England*. Providence, R.I.: Brown University.

__________. 1964. "Interrelation between Regional and Social Dialects." In *Proceedings of the Ninth International Congress of Linguists (1962)*, pp. 135-43. Edited by Horace G. Lunt. The Hague: Mouton & Co.

__________. 1972. *Studies in Area Linguistics*. Bloomington: Indiana University Press.

Labov, William. 1969. "Commentary on Gumperz' "How Can We Describe and Measure the Behavior of Bilingual Groups?" In *Description and Measurement of Bilingualism: An International Seminar (1967)*, pp. 250-5. Edited by L. G. Kelly. Toronto: University of Toronto Press.

__________, ed. 1972a. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

__________, 1972b. "Study of Language in Its Social Context." In *Sociolinguistic Patterns*, pp. 183-259. Edited by William Labov. Philadelphia: University of Pennsylvania Press.

Lehmann, Winfred P. 1962. *Historical Linguistics: An Introduction*. New York: Holt, Rinehart, and Winston.

Lévi-Strauss, Claude. "The Future of Kinship Studies." *Proceedings of the Royal Anthropological Institute of Great Britain and Ireland*, pp. 13-22.

Linguistic Provinces Commission, Report of the Commission. 1948. *Report of the Linguistic Provinces Commission*. New Delhi: Government of India Press.

Lounsbury, Floyd G. 1964. "The Structural Analysis of Kinship Semantics." In *Proceedings of the Ninth International Congress of Linguists*, pp. 1073-90. Edited by Horace G. Lunt. The Hague: Mouton & Co., Reprinted in *CA*, pp. 193-212.

Luce, G. H. 1959. "Introduction to the Comparative Study of Karen Languages." *Journal of the Burma Research Society* 42:1.1-18.

McCormack, William. 1960. "Social Dialects in Dharwar Kannada." In *LDSA*, pp. 79-91.

__________. 1968a. "A Causal Analysis of Caste Dialects." In *SIL*, pp. 221-5.

__________. 1968b. "Occupation and Residence in Relation to Dharwar Dialects." In *SCIS*, pp. 475-85.

MacKenzie, D. N. 1961. "The Origins of Kurdish." *Transactions of the Philological Society of London* 68-76.

__________. 1969. "Iranian Languages." *CTL5*, pp. 450-77.

Mackey, William F. 1965. "Bilingual Interference: Its Analysis and Measurement." *The Journal of Communication* 15:4.239-49.

Macnamara, John. 1969. "How Can one Measure the Extent of a Person's Bilingual Proficiency?" In *Description and Measurement of Bilingualism: An International Seminar (1967)*, pp. 80-97. Edited by L. G. Kelly. Toronto: University of Toronto Press.

Mahadevan, R. 1968. *Gulbarga Kannada: A Brahmin Dialect*. Poona: Deccan College Postgraduate and Research Institute, Linguistic Survey of India Series.

Mahapatra, B. P. 1972. "Ethnolinguistic Approach." *IL* 33:1.24-30.

Martin, Samuel E. 1972. "Nonalphabetic Writing Systems: Some Observations." In *Language by Ear and Eye*, pp. 81-102. Edited by James F. Kavanagh and Ignatius G. Mattingly. Cambridge, Mass.: M. I. T. Press.

Masica, Colin P. 1971. "A Study of the Distribution of Certain Syntactic and Semantic Features in Relation to the Definability of an Indian Linguistic Area." Unpublished Ph.D. dissertation, University of Chicago.

__________. 1974. "The Basic Order Typology as a Definer of an Indian Linguistic Area." In *CCSAL*, pp. 154-80.

Meillet, A. and Cohen, Marcel et al. 1924. *Les Langues du Monde*. Paris: Librairie Ancienne Édouard Champion.

Meenakshisundaran, T. P. 1965. *A History of the Tamil Language*. Poona: Linguistics Society of India, Deccan College, Special Publications No. 2.

Miller, Roy Andrew. 1969. "The Tibeto-Burman Languages of South Asia." *CTL5*, pp. 431-49.

Ministry of Education, Report of the Ministry. 1970. *Report 1967-68.* New Delhi: Government of India Press.

Ministry of Education and Youth Services, Report of the Ministry. 1970. *Report 1969-70.* New Delhi: Government of India Press.

Mishra, R. C. 1971. *hindī kī upabhāṣāe aur dhvaniyā.* Nagpur: *vishvabhāratī prakāshan.*

Morgenstierne, Georg. 1958. "Neu-Iranische Sprachen." *Handbuch der Orientalistik* 4:1.155-77.

Morris-Jones, W. H. 1967. "Language and Region within the Indian Union." In *India and Ceylon: Unity and Diversity*, pp. 67-82. Edited by Philip Mason. London: Oxford University Press.

Müller, Friedrich Max. 1861. *Lectures on the Science of Language.* Delhi: Munshi Ram Manoharlal.

Nayar, Baldev Raj. 1969. *National Communication and Language Policy in India.* New York: Frederick A. Praeger.

Nida, Eugene A. and Fehderau, Harold W. 1970. "Indigenous Pidgins and Koinés." *IJAL* 36.146-155.

Nigam, R. C. 1972. *Language Handbook on Mother Tongues in Census (Census of India, 1971).* New Delhi: Government of India, Census Centenary Monograph No. 10.

Official Language Commission, Report of the Commission. 1956. *Report.* Delhi: Government of India Press.

Official Languages Committee, Report of the Committee. 1960. *Gujarat: Official Languages.* Ahmedabad: Government Press.

O'Neil, Wayne A. 1968. "Transformational Dialectology: Phonology and Syntax." In *Verhandlungen des zweiten internationalen Dialektologenkongresses, 1965*, v. 2, pp. 629-38. Wiesbaden: Franz Steiner Verlag.

Pandit, P. B. 1963. "Sanskritic Clusters and Caste Dialects." *IL* 24. 70-80.

__________. 1965. "Indian Readjustments in the English Consonant System." *IL* 25.202-5.

__________. 1969. "Parameters of Speech Variation in an Indian Community." In *LSIT*, pp. 207-28.

__________. 1972a. "Bilingual's Grammar: Tamil-Saurashtri Grammatical Convergence." In *India as a Socio-Linguistic Area*, pp. 1-25. P. B. Pandit. Ganeshkhind: University of Poona Press.

__________. 1972b. *India as a Socio-Linguistic Area.* (Gune Memorial Lectures, 1968). Gareshkhind: University of Poona.

__________. 1974. "Bilingual's Grammar: A Case Study of Tamil--Saurashtri Number Names." *IJDL* 3:1.181-97.

Pattanayak, D. P. 1975. "Caste and Language." *IJDL* 4:1.97-104.

Pedersen, Holger. 1959. *The Discovery of Language.* Translated by John Webster Spargo. Bloomington & London: Indiana University Press. Original edition published in 1931. [Citations from 1959 edition]

Pieris, Ralph. 1951. "Bilingualism and Cultural Marginality." *British Journal of Sociology* 2.328-39.

Pinnow, Heinz-Jürgen. 1959. *Versuch einer Historischen Lautlehre de Kharia Sprache.* Wiesbaden: Otto Harrassowitz.

__________. 1963. "The Position of the Munda Languages within the Austroasiatic Language Family." In *Linguistic Comparison in South East Asia and the Pacific,* pp. 140-52. Edited by H. L. Shorto. London: School of Oriental and African Studies, University of London.

__________. 1966. "A Comparative Study of the Verb in the Munda Languages." In *Studies in Comparative Austroasiatic Linguistics,* pp. 96-193. Edited by Norman H. Zide. The Hague: Mouton & Co.

Polák, Václav. 1954. "Contributions à l'ètude de notion de langue et de dialecte." *Orbis* 3:1.89-98.

Pyles, Thomas. 1964. *The Origins and Development of the English Language.* New York: Harcourt, Brace, and World.

Ramamurti, G. V. 1931. *A Manual of the Sora Language.* Madras.

Ramanujan, A. K. 1968. "The Structure of Variation: A Study in Caste Dialects." In *SCIS,* pp. 461-74.

__________ and Masica, Colin. 1969. "Toward a Phonological Typology of the Indian Linguistic Area." *CTL5,* pp. 543-77.

Ramaswami (Aiyar) Ayyar, L. V. 1932. "Tulu Prose Texts in Two Dialects." *Bulletin of the School of Oriental Studies, University of London* 6:4. 897-931.

Rao, P. Kodanda. 1969. *Language Issue in the Indian Constituent Assembly (1946-50).* Bombay: The International Book House Pvt. Ltd.

Ray, Amal. 1970. *Tension Areas in India's Federal System.* Calcutta: The World Press Pvt. Ltd.

Report of the Seminar on National Integration. 1961. New Delhi: University Grants Commission.

Robins, R. H. 1967. *A Short History of Linguistics*. Bloomington: Indiana University Press.

Robra, S. K. 1971. "Sindhi Kacchi and Emigrant Sindhi." *IL* 32:2.123-31.

Robson, Barbara. 1975. "Pidgins and Creoles as Natural Languages." Paper presented at the International Conference on Pidgins and Creoles, Honolulu, Hawaii, January 6-10.

Ross, Aileen D. 1965. "Some Social Implications of Multilingualism." In *Towards a Sociology of Culture in India*, pp. 212-36. Edited by T. K. N. Unnithan, Indra Deva, and Yogendra Singh. New Delhi: Prentice Hall of India (Private) Ltd.

Samarin, William J. 1971. "Salient and Substantive Pidginization." In *Pidginization and Creolization of Languages*, pp. 117-40. Edited by Dell Hymes. London: Cambridge University Press.

Sanskrit Commission, Report of the Commission. 1958. *Report 1956-57*. New Delhi: Government of India Press.

Sapir, Edward. 1931. "Conceptual Categories in Primitive Languages." *Science* 74.578. Reprinted in *Language in Culture and Society*, p. 128. Edited by Dell Hymes. New York: Harper & Row, 1964.

Saporta, Sol. 1965. "Ordered Rules, Dialect Differences and Historical Processes." *Lg* 41:2.218-24.

Saussure, Ferdinand de. 1959. *Course in General Linguistics*. Edited by Charles Bally and Albert Sechehaye, in collaboration with Albert Riedlinger. Translated by Wade Baskin. New York: Philosophical Library. Original published as *Cours de linguistique générale*. Paris: Payot, 1931.

Sayce, A. H. 1890. *Introduction to the Science of Language*. London: Kegan Paul, Trench, Trübner and Co., Ltd.

Schiffman, Harold F. 1970a. "Intent Constructions in Dravidian." Paper presented at the Seminar on Regional Universals in Indian Grammar, Berkeley.

__________. 1970b. "Review of *Studies in Indian Linguistics*." *Lg* 46:1. 212-14.

__________. 1973a. "Language, Linguistics, and Politics in Tamilnad." In *Studies in the Language and Culture of South Asia*, pp. 125-34. Edited by Edwin Gerow and Margery D. Lang. Seattle: University of Washington Publications on Asia of the Institute for Comparative and Foreign Area Studies, No. 23.

__________. 1973b. "Language and Dialect." Manuscript.

__________. 1975. "Sociolinguistic Aspects of the Coimbatore Non-Brahman Dialect." Paper presented at the Fourth Conference of the Society for South India Studies, Berkeley, March, 1975.

Schmidt, Johannes. 1872. *Die Verwandtschaftsuerhältnisse der indogermanischen Sprachen*. Weimar: H. Böhlau.

Schuchardt, H. 1889. "Allgemeineres über das Indo-Portugiesische (Asioportugiesische)." *Zeitschrift für Romanische Philologie* 13.476-524.

Secondary Education Commission, Report of the Commission. 1953. "Study of Languages." *Report*. New Delhi: Hindi Union Press, pp. 59-73.

Shafer, Robert. 1941. "Nahali: A Linguistic Study on Paleoethnography." *Harvard Journal of Asiatic Studies* 5:3/4.346-71.

__________. 1949. "Classification of Some Languages of the Himalayas." *Journal of the Bihar Research Society* 36.142-214.

__________. 1952. "Etudes sur L'austroasien." *Bulletin de la Société de Linguistique de Paris* 48:1.111-58.

__________. 1953. "Classification of the Northernmost Naga Languages." *Journal of the Bihar Research Society* 39:3.225-64.

__________. 1955. "Classification of Sino-Tibetan Languages." *Word* 11.94-111.

__________. 1960. "Review of H.-J. Pinnow's *Versuch einer Historischen Lautlehre de Kharia Sprache*." *Word* 16:3.424-33.

Shankara Bhat, D. N. 1960. "Verbal Inflection in Southern Havyaka." *IL* 21.32-40.

__________. 1967-8. "Dialects of Kannada in the Mysore District." *Bulletin of the Deccan College Research Institute* 28:1/2.66-76.

__________. 1968. "The Koraga Language." In *SIL*, pp. 290-5.

Shanmugam Pillai, M. 1960. "Tamil--Literary and Colloquial." In *LDSA*, pp. 27-42.

__________. 1962. "A Tamil Dialect of Ceylon." *IL* 23.90-8.

__________. 1965a. "Caste Isoglosses in Kinship Terms." *AL* 7:3 pt. 2. 59-66.

__________. 1965b. "Merger of Literary and Colloquial Tamil." *AL* 7:4 pt. 2.98-103.

__________. 1968. "Tamil Dialect Notes--Fishermen of Kanyakumari." In *PFICSTS*, vol. 2, pp. 718-28. Reprinted as "Fishermen Tamil of Kanyakumari." *AL* 10:4.1-10. [Citations from *AL*]

__________. 1972. "Address Terms and the Social Hierarchy of the Tamils." In *Proceedings of the First All-India Conference of Dravidian Linguists, 1971*, pp. 424-32. Edited by V. I. Subramoniam. Trivandrum: Dravidian Linguistic Association of India.

__________. 1974. "Code-Switching in the Tamil Novel." In *Structural Approaches to South Indian Studies*, pp. 81-95. Edited by Harry M. Buck and Glenn E. Yocum. Chambersburg, Pa.: Wilson Books.

Sjoberg, Andrée F. 1962. "Coexistant Phonemic Systems in Telugu: A Socio-Cultural Perspective." *Word* 18:3.269-79.

Smith, Ian. "Sri Lanka Portuguese Creole." Unpublished Ph.D. dissertation, Cornell University, forthcoming.

Southworth, Franklin C. 1971. "Detecting Prior Creolization: An Analysis of the Historical Origins of Marathi." In *Pidginization and Creolization of Languages*, pp. 255-76. Edited by Dell Hymes. London: Cambridge University Press.

__________. 1974a. "Linguistic Masks for Power: Some Relationships Between Semantic and Social Change." *AL* 16:5.177-91.

__________. 1974b. "Linguistic Stratigraphy of North India." In *CCSAL*, pp. 201-23.

__________ and Apte, M. L., eds. 1974a. *Contact and Convergence in South Asian Languages*. *IJDL* 3:1.

__________. 1974b. "Introduction." In *CCSAL*.

Southworth, Franklin C. and Daswani, C. J. 1974. *Foundations of Linguistics*. New York: The Free Press.

Spencer, John. 1966. "The Anglo-Indians and their Speech: A Sociolinguistic Essay." *Lingua* 16.57-70.

Sreedhar, M. V. 1964. "A Descriptive Analysis of the Malabari Dialect of Malayalam." Unpublished Ph.D. dissertation, University of Poona.

__________. 1974. *Naga Pidgin: A Sociolinguistic Study of Inter-Lingual Communication Pattern in Nagaland*. Mysore: Central Institute of Indian Languages.

__________. 1975. "Standardization of Naga Pidgin." Paper presented at the International Conference on Pidgins and Creoles, Hawaii.

Srinivas, M. N. 1966. *Social Change in Modern India*. Berkeley and Los Angeles: University of California Press.

Sriramana Acharya, A. 1967. *Halakki Kannada*. Poona: Deccan College Postgraduate and Research Institute, Linguistic Survey of India Series, vol. 1.

Srivastav, R. N. 1965. "K Voprosu o Stratifikatsii Iazyka Khindi." (On the Question of the Stratification of the Hindi Language.) *Vestnik Leningradskogo Universiteta, #8, Seriia Istorii, Iazyka i Literatury, Vypusk 2*. 138-42.

Stampe, David. 1965. "Recent Work in Munda Linguistics: I, II, III, IV." *IJAL* 31.332-41 and 32 (1966) 74-80, 164-8, 390-7.

Stankiewicz, Edward. 1957. "On Discreteness and Continuity in Structural Dialectology." *Word* 13:1.44-59.

Stewart, William A. 1962a. "Creole Languages in the Carribean." In *Study of the Role of Second Languages in Asia, Africa, and Latin America*, pp. 34-53. Edited by Frank A. Rice. Washington, D. C.: Center for Applied Linguistics of the Modern Language Association of America.

__________. 1962b. "An Outline of Linguistic Typology for Describing Multilingualism." In *Study of the Role of Second Languages in Asia, Africa, and Latin America*, pp. 15-25. Edited by Frank A. Rice. Washington, D. C.: Center for Applied Linguistics of the Modern Language Association of America.

Subrahmanyam, P. S. 1968a. *A Descriptive Grammar of Gondi*. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 16.

__________. 1968b. "The Position of Tulu in Dravidian." *IL* (Katre Felicitation Volume, Part 1) 29.47-66.

__________. 1969. "The Central Dravidian Languages." In *Dravidian Linguistics (Proceedings of the Seminar on Comparative Dravidian held at Annamalai University, 1968)*, pp. 107-34. Edited by S. Agesthialingom and N. Kumaraswami Raja. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 17. Reprinted in *Journal of the American Oriental Society* 89:4 (1969) 739-50.

__________. 1971. *Dravidian Verb Morphology (A Comparative Study)*. Annamalainagar: Annamalai University Department of Linguistics, Publication No. 24.

Subramoniam, V. I. 1969. "Malayalam." In *CTL5*, pp. 372-81.

Suseendirarajah, S. 1967. "A Descriptive Study of Ceylon Tamil." Unpublished Ph.D. dissertation, Annamalai University.

__________. 1970. "Reflections of Certain Social Differences in Jaffna Tamil." *AL* 12:7.239-45.

__________. 1973. "Phonology of Sri Lanka Tamil and Indian Tamil Contrasted." *IL* 34:3.171-9.

Tambiah, S. J. 1958. "The Structure of Kinship and its Relationship to Land Possession and Residence in Pata Dumbara, Central Ceylon." *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 88:1.21-44.

__________. 1965. "Kinship Fact and Fiction in Relation to the Kandyan Sinhalese." *Journal of the Royal Anthropological Institute of Great Britain and Ireland* 95:2.131-73.

Taylor, Douglas. 1960. "Language Shift or Changing Relationship?" *IJAL* 26:2.155-61.

Theban, Laurenţiu. 1975. "From Creole Syntax to Universal Semantics." Paper presented at the International Conference on Pidgins and Creoles, Honolulu, Hawaii, January 6-10.

Theban, Maria. 1973. "Structura Propozitiei in Portugheză şi indo-portugheza." *Studii Şi Cercetări Lingvistice,* 5.638-45.

Thompson, R. W. 1961. "A Note on Some Possible Affinities Between the Creole Dialects of the Old World and Those of the New." *Proceedings of the Conference on Creole Language Studies*, pp. 107-13. Edited by R. B. LePage. London: McMillan & Co.

Tiwari, U. 1969. *hindī bhāṣā kā udgam aur vikās.* Allahabad: *līḍar pres.*

Trubetzkoy, N. S. 1958. *Grundzüge der Phonologie.* Göttingen: Vandenhoeck and Ruprecht. *Principes de phonologie.* Translated by J. Cantineau. Paris: C. Klinksieck, 1949. *Osnovy fonologii.* Translated by A. A. Xolodoviča, edited by S. D. Kacnel'sona. Moscow: Foreign Languages Publishing House, 1960. *Principles of Phonology.* Translated by Christiane A. M. Baltake. Berkeley: University of California Press, 1969.

T'sou, Benjamin K. 1975. "The Language of SWONALs (Speakers Without a Native Language): A Study on Accelerated Creolization." Paper presented at the International Conference on Pidgins and Creoles, Hawaii, January 6-10.

Turner, Sir Ralph L. 1966. *A Comparative Dictionary of the Indo-Aryan Languages.* London: Oxford University Press.

Tyler, Stephen A. 1964. "Koya Kinship: The Relation Between Rules and Behavior." Unpublished Ph.D. dissertation, Stanford University.

__________. 1965. "Koya Language Morphology and Patterns of Kinship Behavior." *AA* 67:6 pt. 1.1428-40.

__________. 1966. "Context and Variation in Koya Kinship Terminology." *AA* 68:3.693-707. Reprinted in *CA*, pp. 487-503. Revised and printed as "Context and Alternation in Koya Kinship Terminology." In *Directions in Sociolinguistics: The Ethnography of Communication*, pp. 251-69. Edited by John J. Gumperz and Dell Hymes. New York: Holt, Rinehart and Winston, 1972.

__________. 1968. "Koya Kinship Terminology: The Relation Between Syntactic and Semantic Analysis." *SIL*, pp. 340-59.

__________. 1969a. "Introduction." In *CA*, pp. 1-23. New York: Holt, Rinehart and Winston.

__________. 1969b. *Koya: An Outline Grammar (Gommu Dialect).* Univerisity of California Publications in Linguistics, vol. 54. Berkeley: University of California Press.

Upadhyaya, U. P. 1968. *Nanjangud Kannada: Vakkaliga Dialect.* Poona: Deccan College Postgraduate and Research Institute, Linguistic Survey of India Series, vol. 2.

__________. 1969. "Mapila Malayalam (A Descriptive and Comparative Study)." Unpublished Ph.D. dissertation, University of Poona.

__________. 1971. "Effects of Bilingualism on Bidar Kannada." *IL* 32:2.132-8.

__________. 1972. "Kuruba--A Dravidian Language." In *TSDL*, pp. 307-28.

Uttar Pradesh. 1966. *Facilities Provided for Linguistic Minorities in Uttar Pradesh.* Uttar Pradesh Government: Bhasha Vibhag.

Uttar Pradesh Language Committee, Report of the Committee. 1963. *Uttar Pradesh: Report.* Lucknow: Superintendent, Printing and Stationery.

Vatuk, Sylvia. 1969a. "Reference, Address, and Fictive Kinship in Urban North India." *Ethnology* 8:3.255-72.

__________. 1969b. "A Structural Analysis of the Hindi Kinship Terminology." *Contributions to Indian Sociology* New Series 3.94-115.

Venkatarajulu Chettiyar, V. 1959. "Naṉṉūl." *kalaikkaḷañciyam* ["Encyclopedia"]: vol. 6.319-20. Published by *tamir̤ vaḷarcci kar̤akam* ["Tamil Propagation Society"], Madras: Orient Longmans Ltd.

Vittachi, T. 1958. *Emergency '58: The Story of the Ceylon Race Riots.* London: Andre Deutsch Limited.

Voegelin, C. F., and Harris, Zellig S. 1951. "Methods for Determining Intelligibility Among Dialects of Natural Languages." *PAPS* 95:3.322-9.

Vogt, Hans. 1954. "Language Contracts." *Word* 10:2-3.365-74.

Wallace, Anthony F. C. and Atkins, John. 1960. "The Meaning of Kinship Terms." *AA* 62:1.58-80. Reprinted in *CA*, pp. 345-69.

Weiner, Myron. 1968. *State Politics in India.* New Jersey: Princeton University Press.

Weinreich, Uriel. 1953. *Languages in Contact; Findings and Problems.* New York: Linguistic Circle of New York, Publication No. 1.

__________. 1954. "Is Structural Dialectology Possible." *Word* 10:2/3. 388-400. Reprinted in *RSL*, pp. 305-19. [Citations from *RSL*]

__________. 1957. "Functional Aspects of Indian Bilingualism." *Word* 13:2.203-33.

__________; Labov, William; and Herzog, Marvin I. 1968. "Empirical Foundations for a Theory of Language Change." In *Directions for*

Historical Linguistics: A Symposium, pp. 95-188. Edited by Winfred P. Lehmann and Yakov Malkiel. Austin: University of Texas Press.

Weinreich, Uriel. 1972. *Explorations in Semantic Theory*. The Hague: Mouton & Co.

Whinnom, Keith. 1965. "The Origin of European Based Creoles and Pidgins." *Orbis* 15.509-27.

Wolff, Hans. 1959. "Subsystem Typologies and Area Linguistics." *AL* 1:7.

Wolfram, Walt. 1973. "On What Basis Variable Rules?" In *NWAVE*, pp. 1-12.

__________ and Fasold, Ralph W. 1974. *The Study of Social Dialects in American English*. Englewood Cliffs, New Jersey: Prentice Hall, Inc.

Yadav, R. K. 1966. *The Indian Language Problem*. Delhi: National Publishing House.

Yalman, Nur. 1962. "The Structure of the Sinhalese Kindred: A Reexamination of the Dravidian Terminology." *AA* 64:3 pt. 1.548-75.

__________. 1967. *Under the Bo Tree: Studies in Caste, Kinship and Marriage in the Interior of Ceylon*. Berkeley: University of California Press.

__________. 1969. "The Semantics of Kinship in South India and Ceylon." In *CTL5*, pp. 607-26.

Yesudhason, C. 1975. "Some Sociolinguistic Features of Vilavenko:tu Tamil." Paper presented at the Fifth Seminar on Dravidian Linguistics, Annamalainagar.

Yule, Col. Henry and Burnell, A. C. 1968. *Hobson-Jobson, (A Glossary of Colloquial Anglo-Indian Words and Phrases, and of Kindred Terms, Etymological, Historical, Geographical and Discursive)*, new edition. Edited by William Crooke. New York: Humanities Press, Inc. (Original 1886)

Zide, Norman H. 1969. "Munda and Non-Munda Austroasiatic Languages." *CTL5*, pp. 411-30.

______ ______, and Stampe, D. L. 1968. "The Place of Kharia-Juang in the Munda Family." *SIL*, pp. 370-7.

Zvelebil, Kamil. 1959. "Dialects of Tamil I (Introduction and Texts)." *AO* 27:2.272-317.

__________. 1960. "Dialects of Tamil III: Madurai (Interpretation and Descriptive Statement)." *AO* 28:3.414-56.

__________. 1963a. "Dialects of Tamil IV: Erode, Tuticorin, Ramnad." *AO* 31:4.635-68.

__________. 1963b. "The Vowels of Colloquial Tamil." *AO* 31:2.225-9.

__________. 1964. "Spoken Language of Tamilnad." *AO* 32:2.237-64.

__________. 1966. "Some Features of Ceylon Tamil." *Indo-Iranian Journal* 9:2.113-38.

__________. 1973. *The Irula Language. Neuindische Studien.* Band 2. Wiesbaden: Otto Harrassowitz.

__________ and Viswanathan, S. "Dialects of Tamil II: Madras and Madras Brahmin (Texts)." *AO* 27:4.572-603..